

## Flawed understanding of the scientific process

**Scientific robustness – not certainty – is the key to good environmental regulation. Only when this is recognized will debates about the soundness of scientific advice or access to data be properly resolved.**

**P**ity the US Environmental Protection Agency (EPA). For years, the agency has been attacked for the poor quality of the science on which some of its regulations have been based. Recently, it has been making strenuous efforts to correct such weaknesses, by, for example, setting up a board of scientific counsellors to help reorganize its peer review procedures, and quadrupling its spending on grants to external research teams. But the more EPA seeks to increase its scientific credibility, the more it has come under fire from industrial lobbyists and their allies in Congress.

Recent weeks, for example, have seen the resurrection of earlier proposals to establish a separate National Institute for the Environment (NIE, see page 612). Supporters argue that it is needed to ensure that environmental policy is based — and based only — on high quality science, and that that requires freedom from the intense political pressures which surround the activities of a powerful regulatory agency.

Nothing wrong in that, at least in principle. The problem lies in the questionable motivation and flawed understanding of the scientific process on which such campaigns are frequently based. The revival of interest in the environment institute has been largely prompted by industry's opposition to tough new restrictions proposed by the EPA on particulate emissions (see *Nature* 388, 5; 1997). Despite a large body of epidemiological evidence supporting these proposals on the basis of the thousands of premature deaths they would save, those balking at the costs of their implementation continue to insist that the evidence remains insufficient.

This is where the flawed understanding of science comes in. There

is all the difference in the world between 'good' science, and 'certain' science. Science is continually evolving, and its conclusions are therefore never invalid just because they are provisional or uncertain. It may, for example, be impossible to distinguish a robust correlation between levels of exposure to small particulate matter and the resulting health damage. But that does not mean that there is no correlation, merely that the mechanisms involved remain obscure.

The same suspect motivation and misunderstanding underline current controversies about access to raw scientific data. There have been several recent moves in the US Congress to require federally funded scientists to release the raw data on which their conclusions are based. It is entirely appropriate that, in the appropriate conditions of commercial or personal privacy, access should be granted to those genuinely interested in testing the robustness of conclusions drawn from the data. But in practice this will work only if the difference between robustness and certainty is clearly acknowledged.

The United States does not have a monopoly on such controversies. Similar issues surround both last week's resignation of the chair of a French committee of inquiry set up to test claims of excessive childhood leukaemias around the nuclear reprocessing plant at La Hague, and the difficulties experienced by this inquiry in ensuring the robustness of the local medical data on which conclusions about possible leukaemia clusters have been based (see page 614). The former paid the price of excessive confidence in the safety of nuclear power. Hopefully, those concerned with the latter will carry out their task more open-mindedly. □

## Taking on the barons

**Italy must grasp a unique chance to reform its grant system before the window of opportunity closes.**

**O**ne of the most pressing needs of Italian researchers is for a fair system of peer review, free of the cronyism that continues to dominate the distribution of research grants. Yet in the debate launched by the government's efforts to overhaul Italian research, this issue is as far as ever from a satisfactory conclusion.

As part of a broad assault on inefficiencies in Italy's public sector institutions, a 12-month window has been provided to the Italian government by the so-called Bassanini law to revise the relevant laws by decree. This is a unique opportunity for reform-minded academics and their supporters whose efforts in recent years have been consistently blocked by parliament, many of whose members are university professors keen to defend the privileges of the current system.

Six months down the Bassanini road, the government has put forward proposals for restructuring the way Italian science is run (see pages 609–610). It suggests the creation of new advisory committees within the research ministry, a new pot of strategic funds whose distribution would ultimately be under government control and a system of long-term research planning.

So far, the government's plans remain vague on the details of how

this new system would operate. Yet, as the saying goes, the devil lies in the details. For example, if grant-giving agencies are to involve foreign researchers in the peer review system, they must be able to require that applications are written in English. The Bassanini law could be used to allow this.

There remains strong resistance within the Italian research community to peer review. Decisions on the allocation of grants or academic appointments are traditionally made by committees headed by powerful university professors, known as 'barons', who oppose the loss of power that would accompany a requirement to award on merit rather than personal connection. Their lobbying has blocked the introduction of compulsory peer review for years.

But they should not be allowed to distract the scientific community from exploiting the temporary provisions of the Bassanini law to its advantage. The government should take the details of its research reforms seriously. It would be a serious blow to Italian science if it let itself be distracted, as could easily happen, by the bigger political issues it faces in the autumn, such as consideration of 30,000 tabled amendments to proposed constitutional reforms. □



# Fight over Italian research policy threatens chance for reform

[ROME] As Italy goes into its traditional August torpor, a struggle over who controls national research policy and the allocation of research grants threatens to undermine what could be the country's last opportunity for effective reform of its research structures.

The fighting has been most bitter over the future control, and indeed the very identity, of the National Research Council (CNR), which could see some of its responsibilities for biomedical research re-allocated to a new national institute (see below).

The opportunity for reform was provided by the 'Bassanini law', passed by parliament in March this year. The law, named after minister of public affairs Franco Bassanini, gives the government the right to modify existing laws without formal parliamentary approval in order to improve efficiency in all Italy's public institutions (see *Nature* 386, 208; 1997).

For research, this law opened the possibility of achieving quick solutions to deeply entrenched problems, such as the lack of both long-term research planning and transparent systems of peer review. But conflict between different ministries with interests in research, as well as within the research community itself, has meant that little progress has been made in the past six months—a serious problem, as all decrees prepared under the Bassanini law must be issued before next March.

As a first step in drawing up the new decrees, a working group headed by Giuseppe Tognon, undersecretary of state for research in the ministry of universities and research, catalogued the research activities of the different government-funded research organiza-

tions, and made suggestions as to how they could be more efficiently grouped together.

In addition to the CNR, these agencies include the National Agency for New Technology, Energy and Environment (ENEA), the Italian Space Agency (ASI), the National Institute for Nuclear Physics (INFN) and the various clinical research institutes belonging to the ministry of health.

But in a document presented to the parliament last week, the working group went further than its initial brief to suggest a new system for planning research policy at government level, transferring considerable power from the CNR to the government. Under this proposal, a high-level advisory committee, the Comitato Nazionale per la Ricerca e la Tecnologia (CNRT), would draw up national research priorities. Its nine full-time members, drawn from the scientific and industrial research communities, would be nominated by the prime minister for four-year terms.

The committee would replace the current 25-strong Consiglio Nazionale Scienza Tecnologia (CNST), and, subject to the approval of the government's interministerial financial committee, would control the general distribution of a new interministerial fund intended, among other goals, to support strategic research lines of national interest. This new fund would eventually replace the funds currently allocated to CNR for targeted research.

In addition to the CNRT, eight or nine scientific research councils, to be known collectively as the Consigli Nazionali di Consulenza Scientifica (CNCS), would be established. These would take the place of the



**Bianco: research council needs reform, but must keep its role as "the scientific focus of Italy".**

CNR committees, says Tognon. Their exact function remains undecided, but it will not be confined to CNR issues; for example, they are likely to take on responsibility for evaluating all government-funded research institutes.

Such a restructuring would severely curtail the activities and influence of the CNR. This was established in 1923 as primarily a scientific advisory agency reporting to the head of state and a funding agency for universities, but now embraces a portfolio of 195 research institutes and 137 research centres and groups housed within universities.

The CNR's 14 discipline-based committees have, over time, come to wield what many consider to be an unhealthy amount of power. They are responsible for distributing the CNR's investigator-initiated funds for university research, and nominate research directors of targeted projects and the project committees responsible for their evaluation. ▶

## Government considers plans to create biomedical research institute

[ROME] Another potential threat to Italy's National Research Council (CNR), as politicians consider the future shape of research funding (see above), is a proposal to create a National Institute of Biomedical and Clinical Research.

The ministry of health wants to link its 20 regional clinical research centres into a national institute coordinated by the Rome-based Istituto Superiore di Sanità (ISS), the Higher Institute of Public Health.

This proposal, referred to in the research ministry's document on reform of the Italian research structure, presented to the cabinet last week, could depend

on the outcome of a complete restructuring of the ISS.

In addition to its many research programmes, the 1,500-strong ISS is burdened by many public health responsibilities dealt with by separate ministries.

The Bassanini law (see above) allows the government to rationalize these duties, and the health ministry is now discussing the creation of an independent agency, similar to the US Food and Drug Administration, to take over some of them.

Other ISS duties, such as those related to environmental toxicology, would also be transferred, leaving the ISS as the

central institute of the proposed national network.

Giuseppe Benegiano, the new director of the ISS who took up office in May, opposes these moves. He supports the idea of a national institute but argues that it should retain its regulatory responsibilities. It makes sense financially to keep this within a restructured ISS "because we already have the infrastructure".

Many universities, as well as the research ministry, are concerned about the suggestion of a national institute for biomedical and clinical research for another reason: that it would put responsibility for the financing

of clinical biomedicine firmly in the hands of the health ministry.

The CNR is worried that, once such a national institute is created, there would be political pressure to transfer CNR's own 20-odd biomedical research institutes to the new national body. "Researchers in these institutes want to stay in the CNR," says CNR president, Lucio Bianco.

Giuseppe Tognon, undersecretary of state for research, says that the controversy generated by the plan for a national institute suggests that the issue is likely to be resolved "in the long term rather than short term". **A.A.**

These committees are not obliged to operate a formal peer review system (though some do) or to follow defined rules for evaluation. Their activities are widely criticized as lacking transparency, although many defend them on the grounds that they are democratically elected by the entire academic community.

The heads of the committees form the decision-making Consiglio di presidenza (presidential council) which has considerable influence over the scientific direction of the CNR by choosing the targeted projects and controlling staff appointments. As a result, the CNR president has relatively little personal power.

There have been previous attempts to remove university research funds from the CNR and give them directly to the universities, in a bid to solve the problem of poor distribution. With the removal of its strategic funds, the CNR would lose its role as a funding agency, and be left as merely an organization running research institutes. Tognon argues that the CNR's strength "would remain research". But Lucio Bianco, the agency's new president, says that "it is not correct for CNR just to run institutes". While

acknowledging that CNR needs reform, he thinks this should be an internal matter. "The CNR is the scientific focus of Italy, and it should not be stripped of its powers and responsibilities," he says.

Bianco wants the councils suggested by the ministry to be housed within the CNR, so that they will remain politically independent, and so that the CNR will maintain its "focus" role. He wants the agency role of the CNR to be maintained. But he also wants a clear separation between advisory committees and management functions, and wants to install an appropriate system of peer review.

Bianco insists that the CNR should retain its role as distributor of small grants to universities, but that these should no longer be spread thinly among all applicants. Tognon opposes these grants as "small dowries to each professor", but Bianco argues that they allow the CNR to "scout for new ideas" which would otherwise not find funding.

Bianco and Tognon at least agree on the fact that the CNR's many small research groups and institutes should be rationalized into bigger units. Bianco also wants to establish a new CNR governing body, whose mem-

bers, he suggests, could be in part nominated by the research minister, and in part elected by the academic community. It is important that the CNR president has the casting vote in this body, he says.

Other research organizations have less to fear from the Bassanini reforms. ENEA, for example, is hoping that regulatory reforms will give it more flexibility. When nuclear research ended in Italy after the Chernobyl disaster, ENEA turned its hand to many new research directions — "a thousand flowers bloomed", says its president, Nicola Cabibbo — which it would now like to rationalize.

The CNR still has time to lobby in its defence. Although the decrees which finally emerge from the working document published last week will not require a parliamentary vote, parliament must give an opinion on the directions indicated in the document. Parliamentary discussion will take place in the autumn, and the government is likely to take any strongly supported objection seriously. Discussions may be prolonged and heated, which could threaten the whole point of the exercise — the issuing of decrees within the set time frame.

**Alison Abbott**

## Universities rally against cuts in Pentagon-funded research

[WASHINGTON] A congressional committee has created alarm among many leading US research universities by calling for a major redirection of Pentagon money away from basic research, and for this money to be used to restore the sagging budget for developing and deploying new weapons systems.

The national security appropriations subcommittee in the House of Representatives has cut the Clinton administration's budget request for basic research at the Department of Defense (DoD) from \$1.16 to \$1.03 billion. The subcommittee argues that the military services need the money for weapons modernization.

The bill would transfer most of the money from the basic research budget, which is largely spent in the universities, to exploratory development, which is done in the DoD and industrial laboratories.

News of the proposed cut alarmed universities, which are more heavily dependent on DoD funds than is sometimes realized. The Massachusetts Institute of Technology, for example, gets about \$60 million a year from DoD — one-fifth of all of its research funds, or about the same as it gets from the National Institutes of Health. Predictably it is the engineering and computer science departments that rely most heavily on DoD money.

University representatives meeting in Washington last week agreed to fight the House plan during the congressional recess, in the hope that it will be rejected when conferees from the House and the Senate

finalize a budget bill in September.

George Leventhal, an official at the Association of American Universities (AAU), which represents the 50 leading research universities, says that they hope to get support from defence contractors for the basic research programme.

University representatives deny that the DoD basic research account has received what the subcommittee describes as "never-ending budget growth", pointing out that it has fallen steadily from \$1.4 billion in 1993 to \$1.1 billion this year.

Taking a longer perspective, however, the Pentagon has held its spending power in

basic research at a level of around \$1.1 billion since 1990, during which time its much larger expenditure on research, development, test and evaluation (RDT&E) has slipped in value by one-quarter.

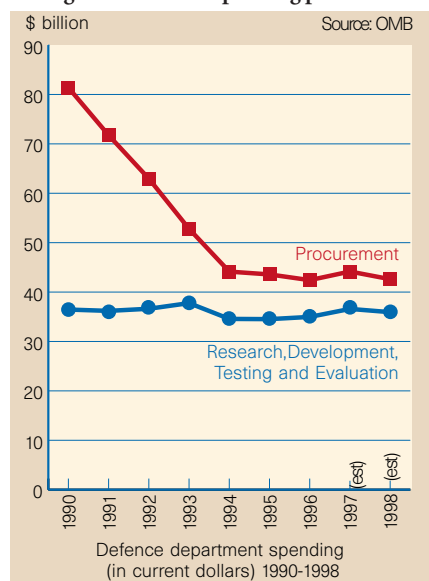
During that period, the Pentagon's procurement of new equipment has nosedived, alarming defence hawks in the Congress. In the 1980s, procurement usually exceeded RDT&E by a factor of three to one. This year, the two items are close to parity — \$36 billion for RDT&E and \$44 billion for procurement of equipment (see diagram).

The Clinton administration had supported a substantial increase in basic research spending this year, arguing that long-term research is a cost-effective way of ensuring the nation's military strength. The Senate appears to accept this, but the House may not.

In a barbed comment aimed at research spending, the language in the House bill complains of cuts in "defence medical programmes, training and readiness accounts, and other programmes such as munitions which have direct and immediate relevance to war-fighting needs".

Universities will try to meet members of Congress, including Bill Young (Republican, Florida), chair of the House appropriations subcommittee, during the recess. They will argue that substantial issues are at stake not just for the universities but for national security, and that the amount of money saved is small compared with procurement and other defence needs.

**Colin Macilwain**





# US to tighten protection of medical data

[WASHINGTON] Biomedical researchers would face federal criminal sanctions for illegally obtaining or divulging protected medical information under recommendations for legislation that Donna Shalala, the US secretary of health and human services, is expected to send to Congress next week.

The recommendations "will make it clear that all researchers must carefully protect the privacy of the personal information they receive — and we recommend penalties if they don't," Shalala said in a speech at the National Press Club in Washington DC.

Declaring that medical privacy is in crisis — a reference to many recent cases concerning the leaking of medical records, such as that of a student in Colorado caught selling patient files to malpractice lawyers — Shalala said it was important that Congress should act rapidly with federal legislation. "We need to enforce our messages with real criminal penalties for abuse."

The Health Insurance Portability and Accountability Act of 1996 requires Shalala to send to Congress by 21 August recommendations on how to ensure the privacy of health information about potentially identifiable individuals. Congress must then write a medical privacy law by 1999; if it fails to do so, she must implement new regulations.

According to sources, Shalala's recommendations will resemble a bill introduced in the House of Representatives by Gary Condit (Democrat, California). This would impose five- to ten-year prison terms on researchers, health-care providers, insurance companies and others with access to private medical information who knowingly disclose or sell it to unauthorized others, or who come into its possession improperly or fraudulently.

In addition to criminal penalties, the Condit bill allows the secretary of health and human services to levy civil fines of up to \$10,000 on those who show a pattern of divulging protected information, a move aimed at those who traffic in medical records rather than those who have merely been careless. It also allows aggrieved patients to sue for compensatory and punitive damages those who unlawfully obtain or divulge information.

The bill defines "protected health information" as "any information, whether oral or recorded in any form or medium... with respect to which there is a reasonable basis to believe that the information can be used to identify the individual".

But, despite the stiff criminal penalties, researchers who behave ethically should not be alarmed if the recommendations become law, says Robert Gellman, an expert on medical privacy policy who helped to draft the Condit bill and headed a subcommittee that advised Shalala on her recommendations.

Gellman says that any law resembling the "middle-of-the-road" Condit bill would be "very simple, straightforward and not threatening" for researchers in practice.

Researchers will have to make sure that they protect information properly, says Gellman. "But it's all going to be administrative and easy to do. The Condit bill gives researchers lots of flexibility." Sharing information between researchers would not be prohibited, provided the researchers have the approval of a local ethics committee known as an institutional review board.

Perhaps most importantly for researchers, both the Condit bill and Shalala's expected recommendations would continue the practice of allowing researchers access to the medical records of identifiable individuals, without consent, provided that the research is determined by a review board to be, as the Condit bill states, "of sufficient importance so as to outweigh the intrusion into the privacy of the protected individual".

In this, the Condit bill is less restrictive than others introduced or being drafted. For instance, a bill introduced in the House by Jim McDermott (Democrat, Washington) would require explicit informed patient consent for research using identifiable records.

Shalala's recommendations are not expected to address the issue of the tens of millions of archived tissue specimens stored in pathology laboratories around the country.

A subcommittee of the National Bioethics Advisory Commission is formulating separate recommendations on this issue.

Nor will the recommendations necessarily be heeded by Congress, which is acutely sensitive to public opinion. In one poll, 85 per cent of Americans called medical privacy very important or essential, and 31 per cent rejected researcher access to records without consent. In Washington, privacy advocacy groups such as the Coalition for Patient Rights are backing more restrictions on researchers.

"Researchers are greatly at risk not because of criminal penalties but because [other proposals have] at least a modest chance of restricting access that they have today," says Gellman.

On the other side of the debate, the pharmaceutical industry has warned against restrictive laws. "Proposals intended to protect individuals from the misuse of medical information could substantially impair our ability to conduct clinical trials," says Richard Kent, vice-president and director of worldwide clinical research at Glaxo-Wellcome.

Addressing a subcommittee of the National Committee on Vital and Health Statistics in February, on behalf of the Pharmaceutical Research and Manufacturers of America, Kent said such proposals would make trials "more expensive or altogether impossible".

Meredith Wadman

## After Dolly, meet Gene, the cloned calf



[LONDON] A US cattle-breeding and biotechnology company last week announced its success in cloning a bull calf, called Gene, using primordial stem cells from a 30-day-old calf fetus. The cloning of Dolly — a transgenic lamb — differed in that the cells were taken from an adult sheep.

ABS Global, based in Wisconsin, also announced plans to set up a new subsidiary,

Infigen, to commercialize the technologies. It claims its process could provide an efficient way of producing large numbers of cloned cattle. The cloning technologies could "offer tremendous promise for enhancing the quality, consistency and nutritional value of dairy and beef products," says Marc van't Noordende, ABS chief executive.

# Pathfinder probes the weather on Mars



[LONDON] After spending more than a month exploring the surface of Mars, the lander Pathfinder is being allowed by the US National Aeronautics and Space Administration to rest at night, in order to cut down on the charging and discharging that depletes its batteries.

The hope is that the lander can keep working for several more months, using solar power. Sojourner is still taking chemical measurements of rocks and soil patches. But over the past couple of weeks, more attention has been focused on the lander's measurements of martian weather. Weather conditions have been found to be steady, with large regular daily swings in pressure and temperature.

But two meteorological oddities have been observed. One is rapid fluctuations in both temperature and pressure. Ground temperatures can change by 20°C within a few minutes, and pressures can also alter over the same timescale. These fluctuations are probably very local, and may be caused by 'dust devils'.

The second oddity is a striking temperature drop within a few feet of the surface. Unlike the Viking landers, Pathfinder can measure temperature and wind speed as a function of height, using thermocouples and

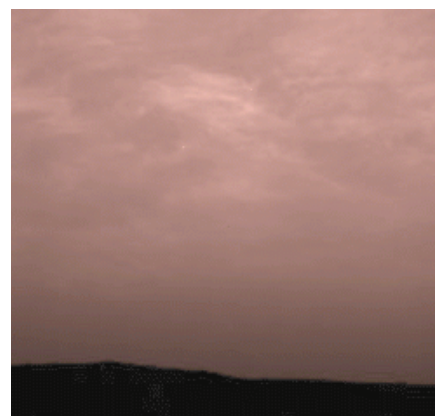
wind socks at different heights on its mast. It has found that the temperature drop from ground level to five feet above the surface can be more than 40°C.

Such information is vital in characterizing the 'boundary-layer' interactions between the martian atmosphere and surface, including factors such as turbulence and rate of heat flow from the surface. Little information has previously been available about such issues, and the findings have significant implications for understanding the behaviour of the atmosphere as a whole, as well as for the way in which wind is involved in erosion and deposition of material on the surface.

Understanding how the winds change, and how they are affected by local terrain and by tides, could be of more than just academic interest. "These local winds will be important if we ever go ballooning on Mars in order to cover substantial distances," says Brian Toon of the University of Colorado.

The dust carried by the winds into the atmosphere also plays an important part in controlling the weather by absorbing heat from the Sun. Two weeks ago, the flight team released pictures of sunrises and sunsets, showing the atmospheric dust.

Some of these pictures also show clouds of water vapour in the early morning — confirmation of the idea that pre-dawn condensa-



**Pink planet:** the martian sky is coloured pink by ice particles formed on dust in early morning.

tion into clouds helps to move water around the planet. In photographs taken in the early morning, the sky is coloured pink by ice that has condensed on dust particles some 16 km above the surface (see above).

The main reason for the stability in the weather conditions experienced by Pathfinder so far is that it has landed during a season of low pressure, when there is minimal atmospheric activity. But the martian autumn will bring dust storms, promising to make the weather more interesting to observers on Earth.

**Stephen Battersby**

## 'Neutral' mechanism sought to fund environmental research

[WASHINGTON] US lawmakers have been left groping for ways to fund scientific research that will not be seen in environmental controversies as either biased or flawed, following the recent bitter debate in Congress about new air pollution standards (see *Nature* 388, 5; 1997).

A committee of the House of Representatives has even dusted off a long-standing proposal to create a National Institute for the Environment (NIE), which could conduct environmental research free of the regulatory agenda that many critics say taints the scientific activities of the Environmental Protection Agency (EPA).

Although the Washington-based Committee for the NIE has long been lobbying for such an institute, the idea has in the past received little political attention. But members of the House appropriations subcommittees that oversee the budgets of the EPA and the National Science Foundation have told the foundation to produce a report by next April on how it might set up and finance such an institute.

Two senior members of the Senate, the Democrat leader, Thomas Daschle of South Dakota, and Chuck Robb (Democrat, Virginia), have gone further, trying —

unsuccessfully — to persuade their colleagues on the appropriations committee to add between \$20 million and \$50 million to the science foundation's budget next year, to enable it to start setting up the NIE.

But, despite this apparent revival in political interest, the concept still has relatively few supporters in Washington, and is unlikely to survive. The White House opposes it, and even many environmental groups are lukewarm. One member of a green lobbying organization calls the NIE "a solution in search of a problem".

Critics of the proposal point out that several federal agencies already conduct peer-reviewed environmental research, and say that creating a new one would waste research dollars on extra administrative overheads. Some question whether it is always desirable to separate scientists from regulators.

They point to the lack of coordination between the Occupational and Health Safety Administration, which drafts regulations to protect workers, and the National Institute for Occupational Safety and Health, which conducts research on job-related hazards. Often, say critics, the two agencies work on unrelated problems, so that the relevant scientific data are not always available when

they are needed for drafting legislation.

But, even if the NIE is shelved, Congress appears anxious to turn to bodies other than the EPA for environmental research. The House appropriations committee, for example, gave a \$35 million increase to the EPA for research on air pollutants, but wants the money transferred to the National Institute of Environmental Health Sciences in North Carolina, part of the National Institutes of Health.

Senate appropriators added less money (\$8 million) for air pollution research, but also specified that it should not go to the EPA. The funds would be used to set up five competitively selected, university-based research centres, which would investigate the health effects of ozone and airborne particulate matter.

Whatever Congress decides, EPA scientists may already be experiencing the political fallout from the defiant stand taken earlier this year by Carol Browner, the administrator of the agency, in favour of tighter air pollution regulations. They can only hope that the controversy will not set back the significant progress made by the agency recently in improving its reputation for high-quality research.

**Tony Reichhardt**



## Independence, but no Nobel winners for India since then

[NEW DELHI] As India celebrates the fiftieth anniversary of its independence today (14 August), many are asking why no Indian scientist has been awarded a Nobel prize in the post-independence era. India's only Nobel winner in science, Chandrasekhara Venkata Raman, was a product of British times; so were Srinivasa Ramanujan, the eminent mathematician, the astrophysicist Meghnad Saha, and Satyendra Bose, of boson fame.

Independent India invested \$50 billion in setting up a huge infrastructure which included about 120 university institutes of technology, and 100 national laboratories, which the country's first prime minister, Jawaharlal Nehru, called "temples of science". But these temples failed to produce gods of the stature of Raman or Bose.

According to leading contemporary Indian scientists, the reason is not difficult to find. "Our system in national labs or universities discourages initiative, innovation or inspiration," says Jayant Narlikar, a prominent astronomer. "It is geared to the progress of the average scientist... as a society, we are uncomfortable with excellence."

Many feel that one mistake India made was to let an excellent university system steadily decay. "Massive reforms are now needed to save these sinking institutions," says Chintamani Nageswara Rao, chairman of the cabinet committee on science and technology. The future of science in India, says Rao, depends on a small percentage of the best young talent who remain working in the few institutions still able to offer good research facilities.

This does not mean that Indian scientists have no reason to rejoice over the golden jubilee of independence. Over the past 50 years, Indian science has eliminated famines, improved life expectancy and brought satellites, supercomputers and atomic power. Food production has increased fourfold, and India is now the world's second-largest milk producer, after the United States.

"But research failed to eradicate poverty or control population growth," says Govindarajan Padmanabhan, director of the Indian Institute of Science in Bangalore.

Rajesh Kochhar, a historian of science, argues that science has failed to make a greater impact because most research in independent India has been carried out with a greater emphasis on achieving personal recognition abroad — a passport to promotion at home — than on its importance in an Indian context. "The Indian nuclear, space and missile programmes constitute science in an extended sense of the term, but are no more than successful applications of known technologies," he says.

**K.S. Jayaraman**

## Make marijuana research easier, panel urges NIH

[SAN FRANCISCO] The US National Institutes of Health (NIH) has been urged by a panel of scientific advisers to set up a centralized mechanism to facilitate the design and conduct of studies of the potential therapeutic effects of marijuana.

In giving its support to such studies, the report, based on a meeting held in Bethesda, Maryland, earlier this year (see *Nature* 385, 756; 1997), argues that science should be separated from the public debate about the potential harm of non-medical marijuana use.

There is no indication yet as to whether the NIH will take up these recommendations. In a statement, Harold Varmus, the director of NIH, said merely that marijuana studies would go through the normal competitive peer review process for funding. "We want to make clear what has always been the case — NIH is open to receiving research grant applications for studies of the medical efficacy of marijuana," said Varmus.

But in principle, the report, released in Washington DC last Friday, could both stimulate further research on marijuana's medical effects and ease the implementation of controversial laws passed last year in California and Arizona that allow doctors to recommend it (see *Nature* 384, 95; 1996).

The California Assembly appropriations committee is currently considering legislation that would create a medical marijuana research programme within the University of California, and allocate \$1 million to its first year of operation.

"This beginning of a thaw on the federal level is a good sign that by the time we're ready to do research, we'll have a partner in NIH, not an enemy," says a senior staffer for California Senator John Vasconcellos (Democrat, San Jose), who introduced the bill.

In its executive summary, the advisory panel asks the NIH to consider administrative mechanisms to encourage grant applications for research into the use of marijuana for appetite stimulation, nausea following anti-cancer therapy, neurological and movement disorders, analgesia and glaucoma.

"For at least some potential indications, marijuana looks promising enough to recommend that there be new controlled studies done," the committee chair, William Beaver, professor of pharmacology and anaesthesia at Georgetown University School of Medicine, says in the report.

The panel was made up of eight experts with backgrounds in clinical studies and therapeutics. According to panel member Reese Jones, professor of psychiatry at the University of California in San Francisco, one reason for the decline in interest in marijuana

research is that the easiest work has already been done, while the more problematic studies have faced a number of hurdles that have made it difficult to pass peer review.

Some of these problem areas include the side-effects of the drug, placebo controls, dose titration, blinding and selection of clinical endpoints. The panel recommends various solutions, for example that the NIH prioritizes the development of a smoke-free inhaled delivery system designed to remove the significant hazards of marijuana's combustion by-products.

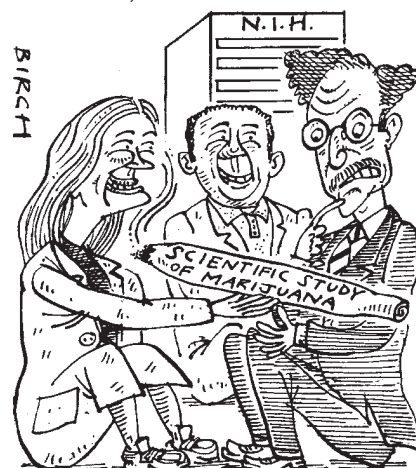
Despite these hurdles, marijuana research should not be ignored, says the panel, adding that there may be compounds in the leaf that are not present in capsule delta-9-tetrahydrocannabinol (THC), the major active ingredient in marijuana smoke, and that the pharmacokinetics and bioavailability of the two are quite different. Some physicians argue that smoking marijuana is more effective than the THC pills in suppressing nausea in cancer patients and stimulating appetite in AIDS patients.

It also pointed out that the difficulties of gaining access to study material have often discouraged researchers. For example, the National Institute on Drug Abuse (NIDA), which grows marijuana for research, has resisted supplying marijuana for therapeutic studies out of fear that it would send the wrong message to the public. The panel recommends that NIH ensures that adequate supplies are available to investigators.

A spokesman for Barry McCaffrey, director of the White House Drug Policy Office, has welcomed the report. "We have said from the start that this is an issue for the medical and scientific community, and not for politics," he told the *Washington Post*.

The full text of the report is available on <http://www.nih.gov/news/medmarijuana/MedicalMarijuana.htm>.

**Sally Lehrman**



# Greenhouse talks edge towards targets

[MUNICH & TOKYO] Little public progress was made towards agreement on binding targets for reducing greenhouse gas emissions during the seventh round of international negotiations on implementing the Rio climate treaty of 1992, which took place in Bonn last week.

But there was some guarded optimism at the end of the meeting that the gap between participants in the discussion may be closing to an extent that such targets can be agreed at the ministerial-level meeting due to take place in Kyoto, Japan, in December.

The US Senate, which would have to ratify any agreement, continues to warn against letting developing countries off the hook. But President Bill Clinton last week told a press conference in Washington that the Kyoto meeting "must take all concrete steps" to address the problem of man-made climate change. "The United States must commit to realistic and binding limits on our emissions of greenhouse gases," Clinton said. "The science demands that we act."

The position of Japan, the host nation for the Kyoto meeting, remains unclear. During the Bonn sessions, Japan's major daily newspaper, *Yomiuri Shinbun*, reported that Japan had proposed informally a strategy to the United States under which all those countries whose emissions are currently below a per capita limit of 3 tonnes would have until 2010 to stabilize per capita emissions at their 1990 levels.

If this proposal were adopted, those countries' absolute emissions — including Japan's — could grow in proportion to their population growth over this period. But countries now above the 3 tonnes limit would have to reduce their emissions to a level at or below 1990 levels in absolute terms by 2010. For the period between 2010 and 2015, governments could choose between either reducing their per capita emissions to under 3 tonnes, or stabilizing emission levels at 1990 levels.

Although it is widely assumed that this proposal represents the position of Japan's powerful Ministry of International Trade and Industry, many feel that it is unlikely to be politically acceptable either in Japan — where there is domestic criticism of the country's relatively weak stand on the climate issue — or with its industrialized partners, which might feel that they were letting Japan off the hook.

Japanese officials have denied the newspaper's report of a firm Japanese proposal, and at a press conference in Bonn the director general of the Japanese Environmental Agency, Ishii Michiko, did not comment on the issue. According to a follow-up report in the newspaper, the German delegation to the Bonn meeting refused to discuss the proposals, saying that it would "not consider anything that has already been reported on in the press".

The final session of the Bonn meeting

heard a potentially significant statement from the Africa group accepting the need for a "globally agreed ceiling" of greenhouse gas emissions, provided that this is based on a convergence on an agreed timescale between reduced emissions from industrialized countries and a "controlled growth" of future emissions from developing countries.

Speaking on behalf of African countries, the Zimbabwean delegate, Rungano Kamanzera, deputy director of the department of meteorological services in Harare, expressed the belief that "this will take us along a path to responsible climate management that allows us to reach our goal of defining a mutually agreed point of convergence and sustainable development".

Last week's sessions also saw detailed discussions about emissions inventory methodologies. The climate convention's Subsidiary Body on Science and Technology proposed that there should be a stronger policy on the scientific aspects of monitoring and reporting emissions. To increase such work, the scientific staff at the United Nations' climate secretariat in Bonn is likely to be doubled, from three to six.

The final preparatory meeting before the Kyoto meeting will take place in Bonn in October. Intense discussions are generally expected at that point, as neither the United States nor Japan is likely to reveal details of their greenhouse-gas reduction targets beforehand. **Quirin Schiermeier & Robert Trindl**

## Head of French leukaemia inquiry quits after partisan remarks

[PARIS] The chairman of a commission of inquiry investigating claims of a slight increase in leukaemias among children living close to the nuclear reprocessing plant at La Hague, on France's northern coast, has resigned following a series of public statements voicing his doubts about the scale of the danger.

The inquiry was set up earlier this year by the ministries of health and of the environment under the former government of Alain Juppé following the publication of the claims in the *British Medical Journal* in January by a team headed by Jean-François Viel of the University of Besançon.

Officials say that the resignation of Charles Souleau, dean of the Faculty of Pharmacy in Chatenay-Malabry outside Paris, is unlikely to have a significant impact on the work of the commission. The inquiry is focusing on two particular aspects: radioecology, under the direction of Annie Sugier of the Institute of Nuclear Protection, and safety and epidemiology, under André Spira of the medical research agency INSERM. Both were nominated by the

government last week.

Souleau's departure is the logical consequence of the public stance he has adopted. In a letter to other members of the committee, he spoke of the "totalitarian vision of the greens", comparing them to "a fundamentalist sect".

During a public meeting in June attended by local councillors from the Nord-Cotentin region in which La Hague is situated, he used conclusions based on a model produced by Cogema, the contractor that operates the reprocessing plant, to argue that the epidemiological risk around the plant was minimal. He challenged the quality of Viel's research, but said Viel had been a victim of an "anglophone lobby" which "controls scientific publications".

Viel says that he does not doubt the intellectual honesty of Souleau, but argues that he has "confused divergences of opinion with strong pressures, which has upset him and pushed him off course". Viel points out that his research was justified by epidemiological analyses carried out in the United Kingdom in 1984 around the

reprocessing facilities at Sellafield and Dounreay. It was therefore logical to carry out a similar study around the French plant.

This view is shared by Sugier, who points out that Viel has not reached any definitive conclusion and that he was merely putting forward a hypothesis on how the leukaemias might be linked to the nuclear activities.

Sugier says it is important to verify both Viel's results and his hypothesis. The real problem appears to be the nature of the data, which are based on the medical registers of the region. These registers are the property of individual doctors, describing how an illness spreads. There are relatively few of them, and they frequently lack rigour in the recording of the data.

The major effort now is to focus on these data and on calculations of dosage based on them. This has the support of Bernard Cazeneuve, parliamentary representative for Cherbourg and chair of a special commission on information about La Hague, which is drafting a bill to be presented to parliament in the autumn on health monitoring in France. **Eric Glover**



# Japan tightens focus in space research

Robert Triendi

**Cuts in Japan's space programme indicate a shift towards users' needs, rather than providing broad-based support for space technology.**

Japan has decided to make major cuts in its space programme, in response to the country's plans for financial reform. A first victim is the space shuttle HOPE, the unmanned space carrier which was intended to function as a supply vehicle to the Japanese module of the International Space Station.

In the longer term, the National Space Development Agency of Japan (NASDA) has indicated that it plans to ensure that its technology development programmes are more immediately responsive to user needs. NASDA may also gradually withdraw from providing the support on which Japan's commercial space industry has grown.

Reductions in NASDA's long-term budget plans were approved last week by the Space Activities Commission, the highest decision-making body on space policy in Japan. This was due to the increasing gap between NASDA's long-term spending plan — agreed at a time when its overall budget was growing at up to 10 per cent a year — and current estimates of its likely annual budget increases over the next few years.

The total budget of the Science and Technology Agency (STA), which oversees NASDA, is likely to rise by 5 per cent next year, in line with the government's commitment to increase funding for science. But the budget increase requested for NASDA has been kept at a mere 1.5 per cent.

NASDA's overall budget has remained almost unchanged since 1994 and, following parliamentary approval of prime minister Ryutaro Hashimoto's financial reform plans, is not expected to rise significantly for several years, according to Mitsugi Chiba, director of STA's space policy division.

## Plans to reduce expenditure

In response, the Space Activities Commission has asked NASDA to plan the implementation of a ¥80 billion (US\$680 million) spending cut in existing projects over the next few years. This amounts to a total of 14 per cent of expenditures for existing projects.

While the spending cuts have been provoked by the need for financial reform, there has been much discussion in recent years about reducing costs. The initial proposal to scrap, scale down or delay some projects — including the abandonment of HOPE — was presented to the commission by NASDA in early July. This has now been approved.

A recent proposal for developing a fully reusable launch vehicle by 2010 is said to have been an important factor in the decision to scrap the HOPE project. On a reduced scale, NASDA will, however, proceed with development of an experimental version, HOPE-X.

The SELENE moon probe mission, a joint undertaking between NASDA and the Institute of Space and Aeronautical Science, Japan's second space agency, is also facing severe cuts. The mission has already been rescheduled, and will now be considerably simplified.

## Low-cost components

Also affected will be NASDA's fledgling satellite development programme. Both the Advanced Land Observation Satellite (ALOS), scheduled for launch in 2002, and the engineering test satellite ETS-VIII, a large satellite intended to test basic space technologies, will have to be built for 20 to 30 per cent less than the initial cost estimates.

NASDA hopes to reduce construction costs of ALOS by reusing components developed for the ADEOS-II remote sensing satellite, a successor to the ADEOS Earth observation platform lost earlier this year (see *Nature* 388, 105; 1997), and by the demise of an engineering design model for ETS-VIII.

Industrial contractors are expected to carry the burden of the cuts. But NASDA has announced that it will assist contractors to develop low-cost satellite components, which should lead to reduced construction costs and projects are said to be underway to develop cost-efficient technologies for satellite components as well as for the Japanese module of the International Space Station.

Eventually, however, NASDA wants companies to develop competitively priced technologies without government support. This appears to be a shift from the space agency's previous policy, under which it had seen a vital part of its mission as being to nurture an internationally competitive space industry.

This reluctance to pursue industrial policy objectives may put NASDA in conflict with the Ministry for International Trade and Industry (MITI) and the Ministry for Posts and Telecommunications (MPT), which want government support for the development of satellite-based telecommunications.

Last April, MPT announced its intention to develop a global multimedia mobile satellite communication system. MPT and



**Vain HOPE?** Japan's space agency has dropped plans to develop the HOPE space carrier.

NASDA are cooperating on a feasibility study to that end. But a similar effort by MITI to launch a programme of low-orbit satellite development was recently described by the Space Activities Commission as "not an appropriate objective of the Japanese space programme".

Faced with the prospect of a declining national space budget, industrial contractors — some of which have announced large investments in novel satellite construction and testing facilities — are seeking new markets and ways of diversifying product lines.

## Need for new markets

NEC, Mitsubishi Electric and Toshiba — the three large satellite manufacturers in Japan — hope that Asian markets will provide them with an entry point into the international commercial satellite business. Japanese contractors are also said to be keen to participate in the emerging national space programmes of neighbouring countries.

Together with the recent ADEOS failure, financial austerity also forces NASDA to shift its technology development strategy towards an increasing recognition of users' needs. Previously, this strategy had been primarily motivated by a desire to reach internationally competitive levels of technological achievement, says Tsukasa Mito, director of the mission-planning division of NASDA's office of satellite systems. In future, he says, greater weight is to be given to the users of the products and services.

As an example, he cites recent consideration of switching the development strategy for environmental monitoring space vehicles from very large platform-type satellites to small low-cost satellites carrying only one or two sensors. This would allow greater flexibility, and the ability to distribute the risk of technological failure across several projects.

But there should also be important benefits to the users of the resulting data. "It is not very practicable to construct large, sophisticated satellites that carry outdated equipment, and deliver their results to clients only with several years delay," says Mito. □

## Tritium leak criticism 'eliminates' contractor from Brookhaven bid

[WASHINGTON] The contractor that for 50 years has managed the Brookhaven National Laboratory on Long Island, New York, will not bid for a new management contract there, its chief announced last week. Lyle Schwartz, president of Associated Universities Inc. (AUI), said that criticism from the US Department of Energy (DOE) over the situation surrounding evidence of tritium leakage has "effectively eliminated" AUI from being a viable bidder in a new bidding round in which DOE is encouraging bids by partnerships.

Federico Peña, the DOE secretary, sacked AUI in May after an internal report criticized management and environmental procedures at the laboratory (see *Nature* 387, 114; 1997). But Schwartz said in a written statement that the department has "focused disproportionate attention on Brookhaven problems and inaccurately and improperly attributed almost all to AUI". This, he said, has scared off potential partners.

Rensselaer Polytechnic Institute of Troy, New York, is joining Westinghouse to bid for the \$400-million contract. The State University of New York at Stony Brook, and Battelle Memorial Institute in Columbus,

Ohio, earlier announced their intention to bid for the Brookhaven contract. Bidding closes on 28 August.

## New minister to reform Spanish universities

[MADRID] A harsh wind of change is expected to blow through Spain's universities in the next few months following a radical switch in leadership at the Ministry of Education and Culture. Fernando Tejerina has been replaced as secretary of state for universities and research by Manuel Jesus Gonzales, a professor at the Open University in Madrid. Tejerina, a professor at the University of Valladolid, was sacked last month by the right-wing education minister, Esperanza Aguirre, because of differences in policy regarding university reforms. Aguirre hopes that Gonzales will support her more hard-line attitude to reform.

## Agreement reached on shorter toxicity tests

[MUNICH] The requirement for long-term toxicity testing of pharmaceuticals on non-rodent animals is to be reduced from 12 to nine months in the United States.

The reduced requirement, which is expected to come into effect next year and which will cut the number of non-rodents

used in toxicity testing by 30 per cent, is the result of a new agreement between the pharmaceutical industries of the United States, Europe and Japan. Europe and Japan require only six months of toxicity data on non-rodents for registration of new pharmaceuticals, and the need for a 12-month testing period in the United States has been widely questioned.

Now, regulatory authorities worldwide are ready to approve new pharmaceuticals with shorter-term toxicity data, according to new consensus guidelines. The guidelines were published last month by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals, and follow a series of prospective and retrospective studies.

## Fresh financial start for Italian space agency

[ROME] The Italian government's economics committee has approved a budget of IL6,500 billion (US\$3.6 billion) for the new five-year plan of the Italian Space Agency. The committee also agreed to wipe out the crippling debts that the agency had accumulated during years of bad management before it was rescued by the appointment of Sergio de Julio as its director last autumn (see *Nature* 386, 311; 1997).

De Julio hopes the space agency will now be able to shed its reputation as an unreliable partner in international collaborations. He has already restored its reputation among Italy's politicians, who had become deeply disenchanted with the space programme following charges of corruption within the agency in the early 1990s.

## Clinicians to tackle close encounters

[BOSTON] About 200 mental health professionals are to attend a 'hallmark conference' this weekend on alien abductions, led by John Mack, a controversial professor of psychiatry at the Harvard Medical School (see *Nature* 376, 457; 1995). The conference, 'Extraordinary Experiences: Clinical Approaches & Cosmological Considerations', will be held at the Omega Institute in Rhinebeck, New York.

The meeting aims to help clinicians address the needs of patients reporting alien abductions and other "anomalous experiences". Clinicians attending, according to the course outline, will receive instruction on diagnosis, countertransference, and insurance reporting. They will also learn how to deal with "possible isolation and stigmatization when working with clients whose experiences do not fit our current understanding of reality".

## Psychology lecturer fired after paedophilia remarks

[EDINBURGH] Chris Brand, a psychology lecturer at the University of Edinburgh, and previously at the centre of a dispute about remarks he is alleged to have made about ethnic background and intelligence, has been dismissed following the unanimous judgement of a university tribunal that he was guilty of gross misconduct over public comments that appeared to condone paedophilia.

The three-person tribunal concluded that the conduct of Chris Brand had been "of a disgraceful nature, incompatible with the duties of (his) office or employment". Brand had been suspended since November after publishing comments in a private Internet newsletter in which, among other things, he queried paedophilia charges against the Nobel laureate Daniel Gajdusek (see *Nature* 387, 117; 1997). Brand says that he intends to appeal against the university's decision.

## US research funds focus on diabetes

[WASHINGTON] The US Department of Health and Human Services is to receive an additional \$150 million over the next five years for research into the prevention and cure of juvenile diabetes. The funding is

contained in the balanced budget act signed into law by President Bill Clinton last week.

Clinton acclaimed the new money as "a tremendous step forward in our fight against diabetes" in a ceremony last Friday (8 August) at the Georgetown University Medical Center in Washington, DC. Another \$150 million will be spent on diabetes research, prevention and treatment among Native Americans, who, for reasons that are at present unknown, suffer disproportionately from diabetes.

## Cassini launch rocket needs leaks repaired

[WASHINGTON] The US Air Force has said it is confident of being able to fix leaks discovered last week during fuelling tests of the upper stage of an unmanned Titan 4B rocket, due to send the probe Cassini to Saturn, in time for the scheduled launch on 6 October.

Officials at the National Aeronautics and Space Administration are concerned that even a slight delay in the launch could reduce the expected scientific payback from the \$3.2-billion, 11-year mission. If the launch were to be delayed beyond mid-November, the changed alignment of the planets, whose gravitational pull is intended to help Cassini to reach its target, could mean that the spacecraft would arrive at Saturn up to several years late.



# A permanent decline in oil production?

**Sir**—The analysis underlying Professor Craig Bond Hatfield's prediction of a permanent decline in world oil production within the next 20 years is questionable both factually and conceptually (*Nature* 387, 121; 1997).

As to the facts, the numbers he gives for present and prospective resources refer only to 'conventional' oil (that is, oil from wells) and ignore other sources such as the Athabasca tar sands and the Orinoco heavy oils. (I understand that Hatfield mentioned them in the original, longer version of his Commentary article.) These alternative sources contain far more oil than can be recovered from wells, but at current crude prices the high cost of extraction has discouraged large-scale production. For the same reason, some conventional oil discoveries have not been developed.

Hatfield's neglect of high-cost oil brings out a conceptual defect of his arithmetic, namely his failure to consider prices. In any mineral, estimates of reserves are meaningful only if they reflect explicit assumptions about prices. When prices change, reserve estimates should be revised. Prices are no less important on the demand side because there is strong evidence of price elasticity in the consumption of energy products. (For a critical survey see D. R. Bohi, *Analyzing Demand Behavior: A Study of Energy Elasticities*, Johns Hopkins Univ. Press, Baltimore, 1981.)

The experience of the 1980s showed that energy use in manufacturing and transportation can be reduced by switching to less energy-intensive equipment and processes.

My interpretation of Hatfield's projections of conventional oil is that, some time between the years 2000 and 2020, crude prices may have to rise to a level where increasing amounts of oil can be profitably extracted from tar sands and the like. This gradual transition should be well

within the power of market forces; futures trading in New York and London enables crude producers, refiners and users to plan several years ahead.

Although the decline in conventional oil output envisaged by Hatfield is indeed possible, total oil output will probably rise for many more decades.

In any case, Hatfield exaggerates the importance of energy in economic growth, which may have been facilitated by low oil prices but is caused primarily by improvements in education, technological change and market-oriented economic policies. These fundamental factors will continue to operate regardless of developments in the oil market.

**Hendrik S. Houthakker**

*Harvard University,  
Cambridge, Massachusetts 02138, USA  
e-mail: hhouthakker@harvard.edu*

**Hatfield replies**—Hendrik S. Houthakker is certainly correct in his assertion that heavy oil and tar sands are huge potential fuel sources, exceeding the world's economically producible conventional oil by a staggering margin. This is also true of oil shale. Such optimistic generalizations, however, do not convey the difficulties involved in large-volume fuel production from these resources.

Tar sands, heavy oil and oil shale did not help us much during the oil shortages of the 1970s, although research on and attempted development of these resources had begun years earlier. Development of economically feasible sources of energy large enough to compensate for the coming decline in petroleum production requires time, gigantic capital investment and, in some cases, innovative technology.

Time is probably the most severe restriction. That was the message in the last sentence of my Commentary article, in which I stressed that society needs to plan

seriously now for the coming decline in the rate of oil production.

I certainly hope that my critics are correct in their view that alternative energy sources will be available to compensate for the decline in conventional oil production, but I fear that they may be too optimistic. During the decline in oil prices in the 1980s, synthetic fuel projects were abandoned as rapidly as they had been initiated during the oil price rises of the 1970s.

Development of substitutes for a significant fraction of our conventional petroleum consumption is a Herculean undertaking that has little chance of success if it can be abandoned in response to short-term changes in the price of oil. The difficulty is not that potential energy sources do not exist. Rather, it is that they are not likely to be online by the time they are needed (2010–2015), considering the current paucity of development effort.

Houthakker also makes the point that shortage of oil is accompanied by higher oil prices, which cause demand to decrease to approach or match supply, thus alleviating or eliminating the shortage. This is an easy way to solve the problem, in that it simply defines shortages out of existence, although oil prices may become enormous and supply quite small. Such a decline in demand during the oil shortages of the 1970s did not prevent high inflation and associated economic hardships.

It is better to recognize the impending reality of genuine shortage and to prepare for it now. If we wait for oil shortages to arrive before we react, as we did in the 1960s and 1970s, we shall once again experience greatly accelerated inflation, a serious reduction in economic growth, and strain on the global monetary system.

**Craig Bond Hatfield**

*Department of Geology,  
University of Toledo, Toledo, Ohio 43606, USA*

## Russian science misrepresented

**Sir**—I am bewildered by Zhores Medvedev's review of *Essays of a Soviet Scientist* by Vitalii I. Gol'danskii (*Nature* 387, 566–567; 1997).

Most of the review is taken up by a derogatory description of the Academy of Sciences of the USSR and of the privileges granted to its members. Only 13 lines are relevant to the subject of the book review. No wonder Medvedev failed to mention the essays "Struggles of a Soviet scientist", "Why

are we lagging behind?", and so on, and in particular the last chapter, which describes the growing anti-semitism in Russia and contains the open letter to President Mikhail Gorbachev on this dangerous trend.

Medvedev asserts that Russian science "has been saved from complete collapse in the past five years only by grants from George Soros and by renting out the buildings of scientific institutes to businesses and foreign companies".

Russian scientists are indeed very grateful for generous financial assistance from Soros (I am myself a Soros Professor),

but the donations of the Soros Foundation to Russian scientists (US\$65 million since 1994) total only just over half as much as the grants from the Russian Foundation for Basic Research (\$122 million).

Meanwhile, the Russian Academy of Sciences received from the federal budget in 1996 (for one year) about \$350 million, far less than is needed.

**Eugen F. Makarov**

*Institute of Chemical Physics  
of the Russian Academy of Sciences,  
Kosygin Street 4,  
Moscow 117334,  
Russia*

# What cannot be said in science

The scientific enterprise is full of experts on specialist areas but woefully short of people with a unified world-view. This state of affairs can only inhibit progress, and could threaten political and financial support for research.

Mott T. Greene

Scientists have an understandable modesty about publicly discussing areas of research other than their own. But this reticence has had the unforeseen consequence that generalization and synthesis, essential parts of the advance of science, are very much neglected. Scientists are trapped in their own specialisms, leaving others, often poorly qualified, to represent to the public the larger architecture and interconnections of modern scientific theories. Although the capacity to convey to society a compelling vision of the whole of science may not be necessary in the day-to-day progress of investigation, it is crucial in maintaining cultural, political and financial support for science.

Scientific education has become so specialized that scientific literacy is little more advanced among scientists than it is among non-scientists. Undergraduates who have completed courses on cell biology and evolution are unable to discuss broad issues in evolutionary theory, let alone Earth history or cosmogony, in any greater depth than can their non-scientist peers. Physics students don't know how a protein differs from a nucleic acid; chemistry students don't know the age of the Earth; geology students cannot give a simple account of metabolism or say why the sky is blue.

This is not to say that science students cannot understand several fields of science and their connections. But a generalized curiosity has not been encouraged or reinforced in basic science training for almost a century. The robust pride that one's knowledge of science is narrow and deep is almost universal among specialists and is powerfully reinforced by three related phenomena.

## Narrow depth of knowledge

First, there is no provision in undergraduate curricula for broad acquaintance with several sciences. The norm in the United States, as in many other places, is self-selection of a single science in the first year. Basic science education is like basic training in the military — directed to tactics rather than to strategy and designed to teach recruits as quickly as possible to use the latest weapons, so that they can be sent to the research front at the earliest opportunity.

Second, there is the problem of the impenetrability of specialist discourse — not only to non-scientists, but to highly trained scientists in different specialisms. Journals such as *Nature* were created to provide rapid publication of results of such importance



MARK DOBSON

that they ought to be communicated beyond the boundaries of individual fields. This function will be frustrated if articles are written in language understood by no-one outside the authors' fields of expertise.

Third, and perhaps more subtle, is the general and strong sense among scientists that, because the advance of science depends on the accumulation of knowledge rather than of opinion, they are not permitted to speak about scientific subjects in public other than those in which they are expert. When Erwin Schrödinger published *What is Life?* in 1944 (ref. 1), he began with an apology: "A scientist is supposed to have a complete and thorough knowledge, at first hand, of some subjects, and therefore is usually expected not to write on any topic of which he is not a master." Freeman Dyson extended this apology in *Origins of Life* more than 40 years later<sup>2</sup>, only to discover that many biologists had not yet forgiven Schrödinger and now were annoyed at him as well.

At less exalted levels of discourse, the imperative to segregate oneself within one's specialism is strong enough to impede the development of interdisciplinary undergraduate science courses even at the first-year level. Most doctoral-level scientists think themselves unqualified to teach a first-year undergraduate course even in a closely allied discipline.

Specialization is not itself the problem.

As the volume of knowledge increases, the proportion of the total comprehended by an individual must diminish. Yet specialization has had unanticipated and even paradoxical consequences.

The paradox is that specialization, however necessary, is not all of science: generalization and synthesis are parts of it as well. Yet generalization and synthesis, even as long as 50 years ago, were well on their way to disappearing altogether from the careers of scientists. I am not referring to textbooks, review articles or the occasional popular lecture, but the deliberate attempt to summarize how the work of one's field fits into the larger framework of scientific advance. Scientists used to do this regularly. That they no longer see synthesis as even a remotely plausible activity is a measure of how completely 'what goes without saying' can pass within a generation or so into 'what cannot be said'.

## Stigma of popularization

The loss of a view of the whole travels in harness with a contempt for generalization — invariably stigmatized as 'popularization' or 'speculation' — and with an irritation directed at those who claim that historically things were different, and even better.

This problem is not resolved even if one takes the position that scientific work is 'self-integrating' — that it is structured so it can function well even if no-one is in charge of

the overall picture. There is still a need for some compelling vision of the whole of science and of its worth to animate those other 'necessary regulators of scientific advance': elected officials who vote on whether to support science, and their constituents.

On the other hand, the unhappiness of many scientists with the picture of the whole presented by historians and others who study science could be simply a negative reaction to seeing the state of science as a whole when it is reasonably well represented, rather than a well-informed reaction to a supposed misrepresentation. One must wonder at the criteria by which scientists can determine the accuracy of any historical representation, given that most have declared themselves ineligible to comment on issues outside their own part of the research front.

But I have never met a scientist, however specialized, who felt she or he could not discuss what science is, or the scientific viewpoint, or the scientific method, or the difference between science and superstition. So an interesting phenomenon has evolved in this century — people who freely admit not knowing much detail about matters beyond their own field of science, and who would never speak about any other aspect of science in particular, but are perfectly comfortable speaking about science in general. How did this happen?

## Birth of subdisciplines

The answer lies not in the fact of specialization, but in the way it comes about. Take a group of investigators working on a problem. They decide that new data of a different kind are needed to solve it. In attempting to obtain this information, they stumble across a genuinely interesting but separate problem. Hence a specialism is born of a failed excursion in which the line of investigation is complex enough to require recruitment of graduate students to finish the job. Students are recruited by presenting the problem as a hot, new, cutting-edge field.

When the students arrive, they hear about 'the problem', which their advisers divide into parts so that no student sees the whole picture. The students graduate, decide that they need an annual meeting (as each of them knows only part of the problem), their own specialist journal, their own students and their own funding — and the sub-field is permanently established.

The first breakaway group was trained in one field, but has helped almost inadvertently to create another. When its members retire and write their memoirs, they recall that the field they helped to start came from another field of which they still feel themselves to be members, and they think of themselves as having helped to solve the original problem. But the students of this first group are the first generation of the new field. And when they retire and write their memoirs, they

recall their good fortune at entering the field "just as their science assumed its modern form". They think of themselves as having supplanted the field of their advisers, in the sense that, even though the field still exists, they see it as behind the current research front. They assume a historical discontinuity — a case of replacement, not of descent.

Hence there are two historical views of the same situation, one recording an evolutionary development and the other a revolutionary rupture. The shift in mental geometry is critical — the first view implies a structure with return as one plausible course of action. But the second view contains no such possibility. This is exactly the moment where 'what goes without saying' turns into 'what cannot be said' — and it happens very, very fast.

There are dozens of examples, but two will suffice. First is the case of cladistic systematists and biogeographers, who see themselves as having supplanted other approaches to systematics and biogeography at the research front. Their technique of exhaustive characterization and splitting based on structural divergence works well with living organisms. But applied to palaeontology, the technique has made it impossible to infer evolutionary descent of very similar but distinct organisms in conformably bedded strata. Cladists are so concerned with the classification of organisms that they have forgotten to tell their students (or were never told themselves) that the point of having a fossil record is to document evolutionary modification and descent. A system of classifying fossils that makes statements about descent impossible is useless for evolutionary biology. It will split biology and palaeontology apart and leave them with two different concepts of a species.

My second example is numerical modelling in the Earth sciences — another new approach that sees itself crowding out older ones at the research front. In constructing numerical models, it is necessary to use real data as a check on the model's adequacy, then to use more and more real data to tweak the model a little better. But there is a danger of using all the data in the world to make the model, to the point that there is nothing left for the model to interpret. Further, there is the danger of accepting existing constraints in model-building as legitimate constraints on theory (taking seriously global climate models with no clouds, for instance).

In both cases, the main aim of the exercise has become to have the 'best model' — of the classification of organisms or some dynamic system in the Earth — without regard to the consequences for the unity of theory or the adequacy of explanations provided for the problems these fields were founded to solve.

But isn't this just how science works? Yes, if 'works' is defined as the current internal standard of self-advance by continued unbalanced specialization and the indefinite postponement of synthesis.

This state of affairs comes about not when scientists are behaving badly, but when they are behaving well; not when they fail to meet the challenges they were trained to face, but when they do meet them. Yet this pattern will not advance scientific literacy, but must further corrode it. It creates cohorts of young scientists who do not perform the integrating function and don't know anyone who does or ever did do it.

There no longer exists a reward structure encouraging scientists in mid-career to pause, reflect and produce careful narratives of continuity of descent, linking new developments to the great structure of science that precedes them, and impartially sorting claims of novelty. No such structure has existed for almost a century.

Historians of science are thus the unintentional residuary legatees of the process of scientific integration and the 'view of the whole'. I should point out, as I am a historian of science, that historians of science could never take over the function of producing general scientific literacy. There are only a few thousand of us in the world, and only a very few study contemporary science in a way requiring complex mastery of the ideas of several fields. There are notable exceptions, for example Steven G. Brush's history of modern planetary physics<sup>3</sup>, but these are not directed at producing general literacy among nonspecialists.

## The point of generalization

In the seventeenth century, people used all the science they knew to explain the operations of the world, and then 'plugged the remaining holes with God'. Now there is enough science for a world-view with widely placed small holes. Yet most people, including increasing numbers of natural scientists, have a world-view with large and rather closely placed holes that they are content to fill either with blithe ignorance or with supernatural explanations for phenomena already well understood in physical terms.

So today's scientists are in command of only a small part of what is known, and there are no educational or career structures that mandate, suggest or reward the synthesis of results into a unified world picture. If this trend continues, one can imagine a world dominated by the results and artefacts of natural science, but in which no-one has a scientific world-view. This outcome, not as bizarre or unlikely as it may appear at first, would be remarkable, not least for the danger it would pose to the continued survival of the scientific enterprise. □

Mott T. Greene is at the University of Puget Sound, Tacoma, Washington 98416, USA.  
e-mail: greene@ups.edu

1. Schrödinger, E. *What is Life?* (Cambridge Univ. Press, 1944).
2. Dyson, F. *Origins of Life* (Cambridge Univ. Press, 1985).
3. Brush, S. *A History of Modern Planetary Physics*. 3 vols (Cambridge Univ. Press, 1996).



# Will solid hydrogen ever be a metal?

Peter P. Edwards and Friedrich Hensel

**At ultra-high pressures, liquid hydrogen becomes metallic. So should solid hydrogen, yet it stubbornly resists. A newly predicted spontaneous asymmetry of molecules in the solid may be the reason.**

In 1926 J. D. Bernal proposed that all matter, when subjected to a high enough pressure, will inevitably become metallic — that is, it will be permeated by a sea of completely free electrons that conduct electricity easily. The most enticing substance for pressure-induced metallization is, in fact, the lightest and supposedly the simplest of all the elements in the periodic table — hydrogen. Wigner and Huntington in 1935 first predicted that molecular diatomic hydrogen would undergo a transition to a metallic state at an imposed pressure of about 250,000 atmospheres<sup>1</sup>; current predictions are in a range close to three million atmospheres. But despite an unrelenting experimental assault at these ultra-high pressures, dense solid hydrogen has so far defied all attempts at metallization<sup>2</sup>. On page 652 of this issue<sup>3</sup>, Edwards and Ashcroft describe a new twist to the continuing saga, which may help explain the recalcitrance of solid hydrogen towards metallization.

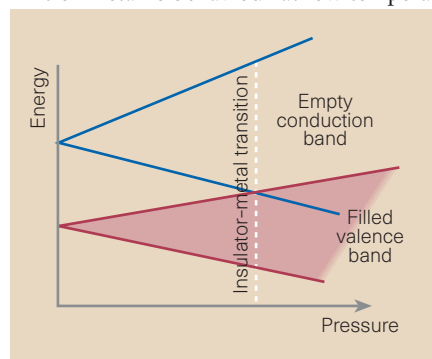
Why are we so interested in making metallic hydrogen? It may have exotic properties and important applications. For example, it has been predicted to be a room-temperature superconductor; and if it could be stably quenched from high pressures to ambient, it would be a highly effective and exceptionally clean propellant for space travel. The metallization of hydrogen would also provide new insights into the nature of giant planets — Jupiter and Saturn, for example, both contain more than 400 Earth masses, most of which is hydrogen under extreme conditions.

Edwards and Ashcroft have discovered a remarkable spontaneous electronic polarization of hydrogen dimer molecules at ultra-high pressures. In stark contrast to the room-pressure situation, in which a hydrogen molecule has precisely equivalent electron (charge) density at both protons, at high pressures the electronic charge piles up preferentially at just one of the constituent protons. The hydrogen molecule thus develops a permanent electric dipole moment. This induced ionic character — in the limit one would regard hydrogen as 'protonium hydride',  $\text{H}^+\text{H}^-$  — may serve to delay, or even thwart completely, the long-awaited transition to the metallic state of dense solid hydrogen.

The pressure-induced metallization of solid hydrogen can be viewed in terms of electronic band theory. At low pressures the solid is an infinite crystalline assembly of isolated  $\text{H}_2$  molecules, all weakly interacting. The electrons are all bound to their molecules, and need to be freed to conduct electricity: in band-theory terms, we have a completely filled valence band separated by a very large electronic energy gap ( $>13$  eV) from the empty conduction band (Fig. 1). Under ambient pressure, elemental solid hydrogen could only conduct by the thermal excitation of large numbers of electrons from the valence band to the conduction band. That would require enormously high temperatures.

Solid hydrogen is thus an insulator. But as the applied pressure increases,  $\text{H}_2$  molecules begin to interact more strongly, and the energy bands widen. In the limit of zero band gap we would have an insulator-to-metal transition — and metallic hydrogen.

Several groups have already succeeded in compressing solid hydrogen to ultra-high pressures, with an elemental density now more than 10 times the room-pressure solid density<sup>2,4</sup>. The average density of electrons in dense solid hydrogen is now two to three times that of the excellent metal, aluminium. And yet, remarkably, solid hydrogen under these crushingly oppressive conditions remains a stubborn insulator, with not even a hint of metallic behaviour at low tempera-



**Figure 1** Bands under pressure. With increasing pressure, and therefore density, the electronic energy bands of solid hydrogen broaden. One assumes that eventually they will overlap, making a conducting, metallic state; but spontaneous polarization of the  $\text{H}_2$  molecules (Fig. 3) might postpone that, or even prohibit it.

tures. The consequences of squeezing solid hydrogen are clearly a good deal more subtle than at first thought, and it turns out that this, the 'simplest' of all molecular solids, has a rich and remarkably complex phase diagram<sup>2,4,5</sup> (Fig. 2).

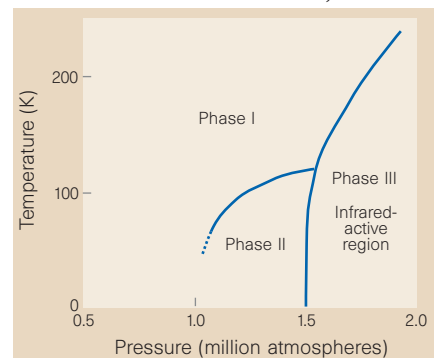
In phase I of solid hydrogen, the so-called orientationally disordered state, the individual molecules execute complete rotational motion in addition to the usual molecular vibrations.

Below a temperature of about 120 K, solid hydrogen undergoes a transition at about 1.5 million atmospheres to phase II, a state in which the constituent  $\text{H}_2$  molecules become 'frozen' in a random orientation in the crystal.

Phase III is undoubtedly the most intriguing, because elemental hydrogen remains in this state up to the highest pressures yet achieved, and still does not become metallic. Perhaps the most remarkable thing about phase III is the appearance of a striking absorption of infrared radiation, which is completely absent in the isolated molecule and largely so in the solid at lower pressures (phases I and II).

Why is the observation of infrared activity in phase III of dense hydrogen particularly important? Substantial infrared activity can only arise from molecules with an intrinsic electric dipole moment. But free  $\text{H}_2$  molecules have a spatially symmetric electron density distribution between their two protons, and so do not possess an intrinsic dipole moment. All hydrogen molecules should therefore show little, if any, infrared activity in the dense solid. However, in 1994 strong infrared activity was detected in solid hydrogen at low temperatures and high pressures, occurring only in phase III. Importantly, this fascinating behaviour occurs under conditions for which solid hydrogen still shows no evidence of metallic behaviour.

Under ambient pressure, states of  $\text{H}_2$  that have any dipolar character are much higher in energy than the unpolarized states. But Edwards and Ashcroft<sup>3</sup> note that, as the band



**Figure 2** A possible phase diagram for dense hydrogen. Phase III, which absorbs infrared light strongly, may be characterized by a spontaneous polarization of the constituent hydrogen molecules.



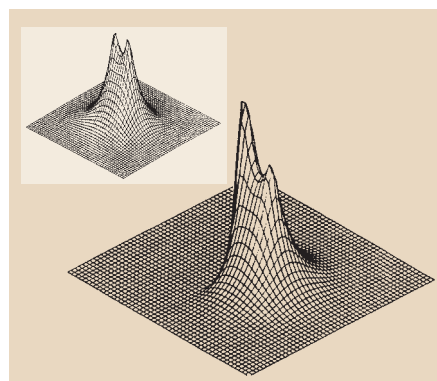


Figure 3 The newly predicted ground state of molecular hydrogen at ultra-high pressure<sup>3</sup>. Here the electron (charge) cloud preferentially accumulates at just one of the two protons, producing an electric dipole. (The magnitude of this charge transfer has been exaggerated for clarity.) The inset shows molecular hydrogen under normal conditions; here the two protons are surrounded by a symmetrical charge cloud.

gap of solid hydrogen closes continuously with pressure, there comes a point at which it may be energetically favourable to mix in a small proportion of these ionic states, resulting in a hybridized ground state. The pressure-induced electric dipole formed on any one of these hydrogen molecules will interact with, and be stabilized by, the other remaining molecules in the dense solid. The stabilization of this new dipolar state will be pressure dependent, and so above a critical density a spontaneous permanent electric polarization sets in on each molecule (Fig. 3).

Edwards and Ashcroft calculate the necessary conditions for the evolution of an induced molecular dipole in hydrogen. They predict the appearance of a spontaneous electronic polarization at a pressure where, experimentally, one sees the onset of the striking infrared activity in hydrogen. So dense hydrogen in phase III is composed of molecular, dipolar hydrogen. There may also be a reorientation and displacement of the dipolar molecules as the new, partly ionic state of hydrogen forms.

But what of the anticipated transition to metallic hydrogen at even higher densities? It has been suggested that molecular hydrogen may eventually become fully ionic, namely  $H^+H^-$ , with enough compression<sup>6</sup>. But from their calculations Edwards and Ashcroft find that the system does not seem to be progressing towards a fully ionic state.

What seems clear, however, is that the presence of even partially ionic character in the ground state of dense solid hydrogen will act to widen the previously narrowing band gap and hence frustrate the transition to the long-sought metallic state. Will solid hydrogen ever become a metal? A pessimistic prospect, contrary to Bernal's optimistic 1926 generalization, might be that solid hydrogen may never achieve metallic status.

Only time — and pressure — will tell. □  
Peter P. Edwards is in the School of Chemistry, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. Friedrich Hensel is in the Fachbereich Chemie, Philipps-Universität Marburg, D-35032 Marburg, Germany.

1. Wigner, E. & Huntington, H. B. *J. Chem. Phys.* **3**, 764–770 (1935).
2. Silvera, I. F. in *Metal–Insulator Transitions Revisited* (eds Edwards, P. P. & Rao, C. N. R.) 21–42 (Taylor & Francis, London, 1995).
3. Edwards, B. & Ashcroft, N. W. *Nature* **388**, 652–655 (1997).
4. Mao, H.-K. & Hemley, R. J. *Rev. Mod. Phys.* **66**, 671–692 (1994).
5. Ashcroft, N. W. *Phys. World* **8**, 43–47 (1995).

## Evolutionary biology

# Even-toed fingerprints on whale ancestry

Michel C. Milinkovitch and J. G. M. Thewissen

Both morphological<sup>1</sup> and molecular<sup>2</sup> studies indicate that cetaceans (whales, dolphins and porpoises) and artiodactyls (even-toed ungulates, which include pigs, hippos, camels and ruminants) form a clade or monophyletic group — that is, they have a common ancestor that is not shared by any other group of mammals. This is counter-intuitive, because it implies that a cow is more closely related to a dolphin or a whale than to a horse, yet it is one of the best examples of congruence between morphological and molecular estimates of mammalian phylogeny.

The molecular analyses of Shimamura *et al.*<sup>3</sup>, reported on page 666 of this issue, further disrupt phylogenetic dogma. Indeed, not only do the authors confirm the close relationship between artiodactyls and cetaceans, but they propose that cetaceans are deeply nested within the phylogenetic tree of the artiodactyls. These results strikingly contradict the common interpretation of the available morphological data (sup-

porting artiodactyl monophyly) and, if correct, would make a cow or a hippopotamus more closely related to a dolphin or a whale than to a pig or a camel (Fig. 1).

Although it is compatible with earlier molecular analyses (for example, refs 4, 5), the idea that cetaceans are highly derived artiodactyls was first suggested in 1994 on the basis of mitochondrial and nuclear DNA and amino-acid sequences<sup>6</sup>. The idea was corroborated by other phylogenetic analyses of DNA sequences<sup>7</sup>. But the issue is still controversial, because the exact means by which molecular sequence data should be analysed remains debated — although analytical settings that are particularly meaningful with respect to phylogenetic inferences can probably be identified in specific instances<sup>8</sup>. But, basically, many morphologists consider that molecular data are necessarily more noisy than morphological data.

The analyses by Shimamura and colleagues now provide a remarkable example of molecular markers, which should lead

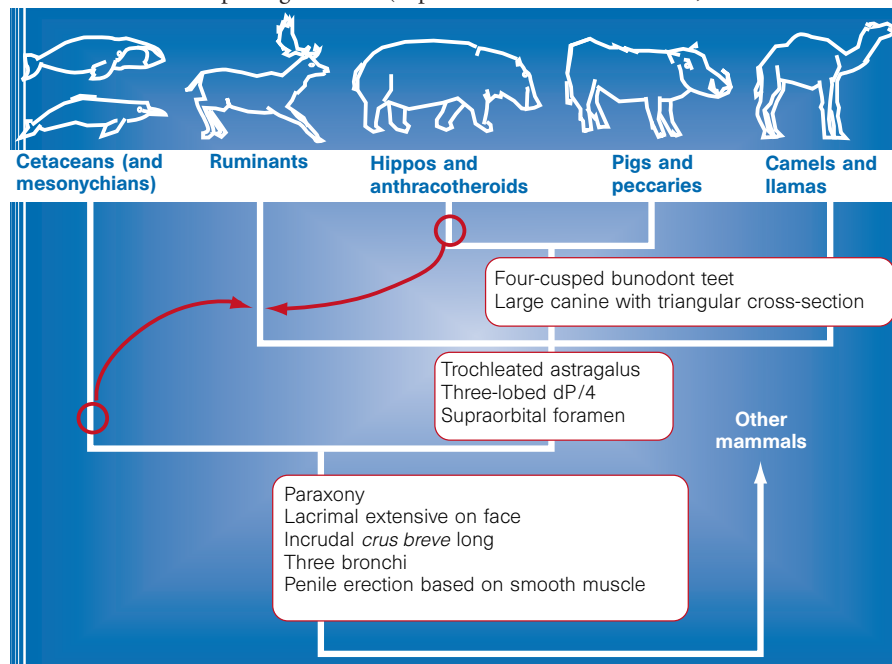


Figure 1 Twisted or untied? Shimamura *et al.*<sup>3</sup> propose to attach the lineages of hippos and cetaceans (curved arrows, red) to the ruminant branch on this phylogenetic tree of artiodactyls — a marked diversion from the traditional view (white branching pattern). The two smaller boxes summarize some of the morphological evidence that disagrees with the new data.

morphologists to re-examine what might have misled them for more than a century. The authors report phylogenetic interpretations of nine retropositional events that led to the insertion of so-called 'short interspersed elements' at particular loci in the nuclear genome of various artiodactyl and cetacean ancestors. Three of these events unambiguously support the grouping of cetaceans, hippos and ruminants in a clade, and the other six provide a partial resolution of the relationships within that clade. Because the likelihood of these elements being independently inserted at the same locus in different lineages (or precisely excised) seems virtually nil, these markers can reasonably be considered to be essentially noise-free.

So, these molecular results may prompt a serious revision of how we view morphological transformations in whales and artiodactyls. Three salient features of artiodactyls are: first, the axis of symmetry of hand and foot runs between the third and fourth digit (paraxony); second, the heel has developed an extremely mobile joint at a place where most mammals have a barely mobile joint<sup>9</sup>; and third, the last lower milk molar consists of three rows of cusps (three-lobed deciduous lower premolar 4, DP/4). Although paraxony is a striking feature, it also occurs in primitive whales<sup>10</sup>, so it is uninformative for the issue at hand. But the remodelled heel is a different matter — because it is present in all artiodactyls and in no other mammal, this character is classically interpreted as derived and, therefore, as supporting artiodactyl monophyly.

The presence of the mobile joint in artiodactyls can be inferred from bones: there is a pulley (or trochlea) on the distal part of the astragalus (one of the heel bones) — a so-called trochleated astragalar head. This led to efficient and fast locomotion in the earliest artiodactyls. Although some features of the artiodactyl heel are present in other mammals such as rabbits, the astragalar head is never trochleated. In modern cetaceans, the hindlimb is so greatly reduced that the heel cannot be recognized, and no complete functional astragalus is known for a fossil whale. However, the mesonychians — a group of land mammals that is considered to be the closest extinct relative of cetaceans — also lack a trochleated astragalus. So if Shimamura and colleagues' hypothesis is correct, then either mesonychians are not closely related to cetaceans (and many dental characters are convergent), or the specialized heel morphology is not the exclusive character that many morphologists take it to be. It may have evolved several times independently in artiodactyls, or have been lost in the mesonychia/cetacean clade. The complete astragalus of an early cetacean would probably shed light on this issue.

The hypothesis put forward by Shima-

mura *et al.* also clashes with the DP/4 character — the tooth has three lobes in all artiodactyls, but not in early cetaceans or mesonychians. If the new molecular data are correct, the morphology of this tooth has, like the trochleated astragalus, a complicated phylogenetic history that includes reversals, convergences or both. Dental differences could reflect dietary differences, because both early cetaceans and mesonychians were probably carrion feeders or carnivores, whereas all early artiodactyls were omnivores or herbivores. But many morphological systematists are reluctant to let functional arguments influence their evaluation of characters.

What characters, besides the molecular ones described by Shimamura *et al.* and by Gates<sup>7</sup>, do hippos, ruminants and cetaceans have in common that makes them different from pigs, peccaries and tylopods (camels and llamas)? Morphological studies have usually upheld close genealogical ties among pigs, peccaries and hippos (and the larger group that includes hippos: anthracotheroids), but the similarities are usually uninformative primitive characters that are also present in the ancestral artiodactyl, or features that are subject to rampant convergences (such as certain dental characters, see Fig. 1). Shimamura *et al.* also suggest that artiodactyls and cetaceans diverged in the Cretaceous — about 15 million years before they are found in the fossil record. However, if hippos are closely related to cetaceans, and if mesonychians are not, then the discrepancy in the times of origin is considerably less: it amounts to the difference between the origin of cetaceans (approximately 50 million years ago) and that of anthracotheroids (around 49 million years ago). But because of the rapid and unique specialization of cetacean morphology, few characters can be recruited to bolster the grouping of ruminants, hippos and cetaceans into a clade. Recovery and study of the earliest cetaceans would probably help to resolve this problem.

Few molecular studies rule out previously accepted morphological trees as convincingly as that of Shimamura *et al.* However, the face of phylogenetic science itself is changing rapidly — it is becoming more objective and less inductive. For example, phylogeneticists no longer need to ask whether molecular data are superior or inferior to morphological data, because the signal-to-noise ratio in morphological, molecular and combined data sets can now be measured directly, even before examining phylogenetic trees<sup>11</sup>. In any case, the new analyses indicate that the use of retropositional events as molecular markers may define a new power of resolution in estimating phylogenies. This method could even bring to a close some of the most intense controversies<sup>2,12,13</sup> in the field, such as whether the toothed whales<sup>2,12</sup> are monophyletic or paraphyletic, and likewise for the rodents<sup>13</sup>. □



#### 100 YEARS AGO

'Cyclone sail' — I have sent to you, for publication, if you think desirable, a photograph of a type of an ideal sail — ideal, in that the wind acting on it has no tendency whatever to incline the boat. The wind pressure acts practically at right angles to the mean surface of the sail. When the wind is making a large angle with the sail, the centre of pressure is almost at the centre of the surface, but when the wind strikes the sail at an acute angle, as in all sails or kites, the centre of pressure moves towards the weather edge; but by suitably adjusting the sail, the desirable result of obliterating all heeling movement has been achieved.

Percy S. Pilcher

From *Nature* 12 August 1897.



#### 50 YEARS AGO

On June 6, 1942, there was a second meeting in Berlin, when the results of the uranium project were reported to Speer, as Minister for War Production. The facts reported were as follows: definite proof had been obtained that the technical utilization of atomic energy in a uranium pile was possible. Moreover, it was to be expected on theoretical grounds that an explosive for atomic bombs could be produced in such a pile.... Following this meeting, which was decisive for the project, Speer ruled that the work was to go forward as before on a comparatively small scale. Thus the only goal attainable was the development of a uranium pile producing energy as a prime mover — in fact, future work was directed entirely towards this one aim.

Prof. W. Heisenberg

From *Nature* 16 August 1947.

Many more abstracts like these can be found in *A Beside Nature: Genius and Eccentricity in Science, 1869–1953*, edited by Walter Gratzer. Contact David Plant. e-mail: subscriptions@nature.com

Michel C. Milinkovitch is in the Unit of Evolutionary Genetics, Department of Molecular Biology, Free University of Brussels, cp 244, Boulevard du triomphe, 1050 Brussels, Belgium.

J. G. M. Thewissen is in the Department of Anatomy, Northeastern Ohio Universities, College of Medicine, PO Box 95, Rootstown, Ohio 44272-0095, USA.

1. Thewissen, J. G. M. & Hussain, S. T. *Nature* **361**, 444–445 (1993).
2. Milinkovitch, M. C. *Trends Ecol. Evol.* **10**, 328–334 (1995).
3. Shimamura, M. *et al. Nature* **388**, 666–670 (1997).
4. Irwin, D. M., Kocher, T. D. & Wilson, A. C. *J. Mol. Evol.* **32**, 128–144 (1991).

5. Czelusniak, J. *et al. Current Mammalogy* Vol. 2 (ed. Genoways, H. H.) 545–572 (Plenum, New York, 1990).
6. Graur, D. & Higgins, D. G. *Mol. Biol. Evol.* **11**, 357–364 (1994).
7. Gatesy, J. *Mol. Biol. Evol.* **14**, 537–543 (1997).
8. Milinkovitch, M. C. *et al. Genetics* **144**, 1817–1833 (1996).
9. Schaeffer, B. *Am. Mus. Novit.* **1356**, 1–24 (1947).
10. Thewissen, J. G. M., Madar, S. I. & Hussain, S. T. *Cour. Forschungsinst Senckenbg* **191**, 1–86 (1996).
11. Lyons-Weiler, J., Hoelzer, G. A. & Tausch, R. J. *Mol. Biol. Evol.* **13**, 749–757 (1996).
12. Milinkovitch, M. C., Orti, G. & Meyer, A. *Nature* **361**, 346–348 (1993).
13. Hasegawa, M., Cao, Y., Adachi, J. & Taka-aki, Y. *Nature* **355**, 595 (1992).

## Adenosine receptors

# Knockouts anxious for new therapy

Solomon H. Snyder

Over the past two decades, putative neurotransmitters have proliferated, from biogenic amines to amino acids, neuropeptides and, most recently, gases such as nitric oxide and carbon monoxide. Amidst the ferment over recently discovered messengers, scant attention has been paid to an atypical neurotransmitter whose candidacy dates back almost 30 years — adenosine. But the field may now be rejuvenated, thanks to a report by Ledent *et al.*<sup>1</sup> (page 674 of this issue) showing that there are behavioural and physiological alterations in mice lacking one adenosine-receptor subtype,  $A_{2a}$ .

Neurotransmission by adenosine and ATP — designated 'purinergic' by Burnstock<sup>2</sup> — is difficult to study because both molecules are involved in many pathways of general cellular metabolism. Nonetheless, the evidence for adenosine as a neurotransmitter is as persuasive as for any other substance: immunohistochemical studies have shown that adenosine is found in discrete neuronal populations in which those cells that use it as a neurotransmitter have much higher localized densities<sup>3</sup>; an efficient, energy-requiring uptake system exists which could account for synaptic inactivation of adenosine; and adenosine has prominent functions outside the brain. Studies by Berne<sup>4</sup> have shown that oxygen starvation leads to a massive accumulation of adenosine, which causes vasodilation and increased blood flow.

The most compelling evidence for the specific actions of adenosine has come from the characterization of its receptors. The influences of adenosine on levels of cyclic AMP in brain slices provided the first molecular evidence for the function of receptors

and differentiation of subtypes. At nanomolar concentrations, adenosine lowers the activity of adenylyl cyclase (which converts ATP to cAMP) through adenosine  $A_1$  receptors, whereas micromolar levels augment adenylyl-cyclase activity at  $A_2$  receptors<sup>5</sup>.

Xanthines and caffeine both block adenosine receptors<sup>6</sup>, and this ability could account for their effects as behavioural stimulants. Xanthine derivatives can block adenosine receptors in direct proportion to their potencies as behavioural stimulants, and at concentrations similar to blood levels of caffeine after a few cups of coffee<sup>7,8</sup>. The bronchodilating, antiasthmatic effects of theophylline and other xanthines may involve adenosine-receptor blockade, although the evidence is controversial.

Molecular-biological techniques have further discriminated receptor subtypes, specifically  $A_1$ ,  $A_{2a}$ ,  $A_{2b}$  and  $A_3$  (ref. 9). The  $A_{2a}$  receptor (which stimulates adenylyl cyclase) is particularly abundant in the basal ganglia, blood vessels and platelets, so it was selected by Ledent *et al.*<sup>1</sup> for targeted deletion.

The  $A_{2a}$ -receptor knockout mice seem to be normal on a gross level, and they breed successfully. But whereas the  $A_2$ -agonist drug CGS-21680 reduces locomotor activity in wild-type mice, it has no behavioural effect in the knockouts. Strikingly, caffeine, which normally stimulates locomotor activity, substantially depresses activity in the knockout mice. These findings cement the conclusion that the stimulant effects of caffeine are derived from adenosine-receptor blockade. But why should caffeine depress behaviour in the knockouts? Adenosine analogues and caffeine can either stimulate or depress locomotor activity

depending on the dose<sup>7</sup>. So, in the knockouts, a stimulatory action of adenosine — perhaps through  $A_1$  receptors — may be unmasked.

Ledent *et al.* also found that the knockout mice seem to be more anxious and aggressive than wild-type animals. This fits with abundant clinical evidence that caffeine increases anxiety, and that stimulation of the adenosine receptor can relieve anxiety. Adenosine modulates pain pathways in complex ways. There is evidence that stimulation of  $A_{2a}$  receptors on sensory pain fibres increases pain perception. The knockout mice show reduced pain responses, suggesting that — at least at  $A_2$  receptors — adenosine normally augments pain perception. This also fits with observations that caffeine can be analgesic.

Adenosine has many effects in the cardiovascular system; it inhibits platelet aggregation and dilates several vascular beds. The knockout mice show more efficient platelet aggregation than wild-type animals, and they are resistant to inhibition of platelet aggregation by adenosine analogues. Ledent *et al.* also found that the knockout mice have increased blood pressure, supporting a tonic vasodilatory role for adenosine.

Pinning down the diverse biological functions of  $A_{2a}$  receptors should provide a route for the pharmaceutical industry to develop related therapeutic agents (Table 1). One possibility might be new cognitive stimulants. Although the relative merits of caffeine's stimulant effects are a source for debate, its cognition-enhancing actions are well documented. Agents that elicit these beneficial influences, without side-effects such as tachycardia, increased urination and agitation, may treat cognitive impairment of the elderly. At present, despite the wealth of available analgesics, none is entirely satisfactory. Moreover, the complex effects of adenosine on pain perception have thwarted efforts to develop adenosine-based analgesics<sup>10</sup>. The findings by Ledent *et al.*<sup>1</sup> from the knockout mice favour selective  $A_{2a}$  antagonists as non-addictive analgesics. Finally, antagonists to the  $A_{2a}$  receptor might improve defects in blood clotting and vascular shock, whereas  $A_2$  agonists should decrease clotting and guard against low blood pressure. □

Solomon H. Snyder is in the Department of Neuroscience, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.

Table 1 Potential adenosine-receptor drug development based on the  $A_{2a}$  knockout phenotype

| Phenotype                              | Drug                                                                        |
|----------------------------------------|-----------------------------------------------------------------------------|
| Altered locomotor response to caffeine | $A_2$ antagonists as cognition enhancers                                    |
| Apparent anxiety                       | $A_2$ agonists as anxiolytics                                               |
| Lesser pain responses                  | $A_2$ antagonists as analgesics                                             |
| Increased platelet aggregation         | $A_2$ agonists as anticoagulants;<br>$A_2$ antagonists for clotting defects |
| Hypertension                           | $A_2$ agonists as antihypertensives;<br>$A_2$ antagonists for shock         |

1. Ledent, C. *et al. Nature* **388**, 674–678 (1997).
2. Burnstock, G. *J. Physiol. (Lond.)* **313**, 1–35 (1981).
3. Braas, K. M., Newby, A. C., Wilson, V. S. & Snyder, S. H. *J. Neurosci.* **6**, 1952–1961 (1986).
4. Berne, R. M. *Circ. Res.* **46**, 807–813 (1980).
5. Daly, J. W. *Cyclic Nucleotides in the Nervous System* (Plenum, New York, 1977).
6. Satin, A. & Rall, T. W. *Mol. Pharmacol.* **6**, 13–23 (1970).
7. Katims, J. J., Annau, Z. & Snyder, S. H. *J. Pharmacol. Exp. Ther.* **227**, 167–173 (1983).
8. Fredholm, B. B. *Pharmacol. Toxicol.* **76**, 93–101 (1995).
9. Olah, M. E. & Stiles, G. L. *Annu. Rev. Pharmacol. Toxicol.* **35**, 581–606 (1995).
10. Sawynok, J. & Yaksh, T. L. *Pharmacol. Rev.* **45**, 43–85 (1993).



## Atmospheric chemistry

## Ozone clouds over the Atlantic

Paul Crutzen and Mark Lawrence

Ozone in the atmosphere has many functions. Especially important to us is that it acts as a filter against solar ultraviolet radiation, protecting the biosphere from a large fraction of the biologically active and potentially harmful ultraviolet radiation of wavelengths less than about 325 nm. About 90% of all ozone is located in the stratosphere (above 10 to 20 km) and only 10% in the troposphere (the region down to ground level). The two regions are separated, dynamically, by the tropopause, but, as reported by Suhre *et al.* on page 661 of this issue<sup>1</sup>, new measurements taken on airliners show that, contrary to earlier information, pockets of extremely high ozone concentration are somehow getting from the stratosphere into the troposphere over the tropical Atlantic.

Both regions are substantially affected by human activities. Ozone concentrations in the troposphere have increased, unlike those in the stratosphere. This is obvious during local episodes of photochemical smog near the ground, but also occurs in wider regions that are affected by anthropogenic emissions of hydrocarbons, carbon monoxide (CO) and nitric oxide (NO), such as the mid-latitude industrial zone of the Northern Hemisphere and the continental tropics and subtropics, which are strongly polluted by biomass burning during the dry season (Fig. 1). In high concentrations ozone can be deleterious to the biosphere, affecting human health, plant growth and agricultural productivity (ref. 2 and refs therein).

The effects of ozone and ultraviolet radiation in the troposphere are, however, not only negative. They produce hydroxyl (OH) radicals, which are involved in the removal of almost all atmospheric gases emitted by human and natural activities — despite very low tropospheric concentrations, about 4 in  $10^{14}$  molecules of air, hydroxyl's high reactivity makes it the detergent of the atmosphere. Most hydroxyl radicals are formed when ozone absorbs solar ultraviolet (UV-B) radiation of wavelengths less than 325 nm. This leads to the production of electronically excited oxygen atoms which have enough energy to react with water vapour to produce hydroxyl radicals.

There is less total overhead ozone in the tropics than anywhere else, so more UV-B radiation penetrates. There is also more water in the tropical atmosphere, so the production of hydroxyl radicals is higher there than elsewhere. A quantitative understanding of the chemistry of the global atmosphere requires, therefore, good knowledge of the chemistry of the tropics and subtropics. This is all the more important as the greatest

impact of human activities on atmospheric chemistry is probably going to occur in these parts of the world.

Unfortunately, information about ozone and other important chemical species is very much lacking in these regions. Regular vertical profiles of ozone are measured at only a handful of tropical stations. Past measurement campaigns have shown a high variability in space and time in the tropics and subtropics. For example, high concentrations are measured over large parts of the rural areas of South America, Africa and Southeast Asia during the dry season, when biomass burning releases a chemical mixture that is the precursor of photochemical smog (CO, hydrocarbons and NO<sub>x</sub>). Between 2 and  $5 \times 10^{12}$  kg of biomass carbon are burned each year in the tropical and subtropical world<sup>3</sup>. This is dirty burning without emission controls. In comparison, global fossil fuel burning is about  $5.5 \times 10^{12}$  kg a year.

In contrast, extremely low ozone concentrations have been measured in the cleanest region of our globe, that is, over the tropical Pacific. Vertical ozone soundings carried out from a ship between the Solomon Islands and Christmas Island<sup>4</sup> showed very low ozone volume mixing ratios of less than 10 p.p.b. (parts per billion) within a few kilometres of the sea surface. That was expected, owing to strong ozone loss by hydroxyl-forming reactions. Unlike measurements made anywhere else in the world, however, there was sometimes almost zero ozone concentrations in the upper troposphere. These low-ozone events appear to be related to intense convective activity and humidification of the air by evaporation of ice particles lifted up by the convection. Higher ozone values were generally

found in drier air in the upper troposphere, between 10 and 12 km, indicating some downward transfer of ozone from the dry stratosphere. But O<sub>3</sub> ratios remained below 30 p.p.b.

Now for the next great surprise in the ozone saga: over the tropical Atlantic, between 10 and 12 km, Suhre *et al.*<sup>1</sup> found extremely high ozone concentrations of up to 500 p.p.b., together with high humidities, in convective regions 5–80 km across. These observations were made by automated instrumentation on an ordinary passenger aircraft. How can this be explained?

Leaving aside the unlikely explanation that lightning causes these very high ozone peaks, they must be due to transport from the stratosphere. There are two possibilities: downward turbulent transfer through the tropopause, caused by convection; or advection of air masses from outside the tropics through a break in the tropopause that is known to exist in the subtropics. Contrary to the Pacific measurements, the high ozone values were not accompanied by low water concentrations (which is a characteristic of stratospheric air). So the air must have been humidified, possibly through the evaporation of ice particles falling from the anvils of thunderclouds, accompanied by downward mixing of the ozone-rich air parcels.

Why this happens over the Atlantic, yet was not observed over the convective part of the tropical Pacific, requires further exploration. Unfortunately, the chemical instrumentation on the MOZAIC aircraft is restricted to O<sub>3</sub> and H<sub>2</sub>O. It would have been useful to have measurements of other chemical constituents, such as CO, CO<sub>2</sub>, N<sub>2</sub>O and halocarbons: if the high-ozone pockets come from the extra-tropical stratosphere, the air would have been chemically processed during its journey, leading to lower amounts of N<sub>2</sub>O and halocarbons, and higher amounts of more reactive nitrogen and halogen oxides — with potentially interesting chemical behaviour in the presence of ice. ►

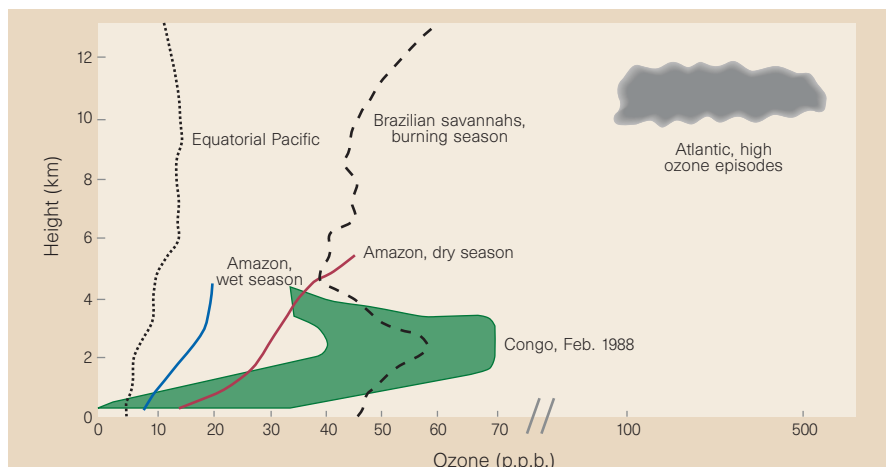


Figure 1 Ozone profiles in the tropics show the contrasts between dry and wet seasons, and between continental and marine sites. Episodes of extremely high ozone concentration, now seen over the Atlantic, are unlike anything reported before. These air pockets must somehow be getting into the troposphere from the stratosphere above.

The impact of the high ozone intrusions on the overall photochemistry of the tropical troposphere cannot be estimated from the data of Suhre et al. The authors have, however, clearly demonstrated the great opportunities and cost effectiveness offered by instrumentation on commercial aircraft. The results show again very clearly how little we know about tropospheric ozone in the tropics and subtropics, which are so important photochemically. Until these regions are much more intensely researched, the quanti-

tative basis for chemical studies will remain weak. The participation of scientists from the tropics and subtropics in this research should be encouraged. □

Paul Crutzen and Mark Lawrence are in the Max-Planck-Institut für Chemie, Abteilung Chemie der Atmosphäre, Postfach 3060, 55020 Mainz, Germany.

1. Suhre, K. et al. *Nature* **388**, 661–663 (1997).
2. Chameides, W. L., Kasibhatla, P. S., Yienger, J. & Levy, H. *Science* **264**, 74–77 (1994).
3. Crutzen, P. J. & Andreae, M. O. *Science* **250**, 1669–1678 (1990).
4. Kley, D. et al. *Science* **274**, 230–233 (1996).

## Immunotherapy

# Fusion induces tumour rejection

Ian Hart and Camilo Colaco

The war on cancer has been waged with ever-increasing intensity over the past 30 years or so, using the weapons of surgery, radiotherapy and chemotherapy. But it has resulted, to date, in only limited advances. This has led to calls for the main directions of research to be realigned, from treatment to prevention<sup>1</sup>. But cancer therapy will be given a renewed impetus by a paper published in *Nature Medicine* by Kufe and co-workers<sup>2</sup>, who describe an elegant but simple strategy for the treatment of cancer using immunotherapy.

The idea of immunotherapy as a new approach may seem to be a contradiction in terms, since strategies based on this concept have been applied clinically for over 100 years, with limited success. Nonetheless, there is renewed optimism<sup>3,4</sup>, largely based on a better understanding of how our cytotoxic T lymphocytes (CTLs) respond to tumour cells<sup>5</sup> (Fig. 1). Central to the induction of CTL responses are professional antigen-presenting cells. These include the B lymphocytes, mononuclear phagocytes and dendritic cells, which present antigen to the immune system in the context of major histocompatibility complex (MHC) class I and class II molecules<sup>6</sup>.

In a host of experiments over the past few years, dendritic cells have been used to process and present tumour antigens. This has generally been achieved by pulsing cultured dendritic cells with peptides eluted from class I MHC molecules<sup>7</sup>, tumour cell membranes<sup>7</sup> or RNA derived from the neoplastic cells<sup>8</sup>. Kufe and co-workers have pursued the alternative strategy of introducing tumour antigens into dendritic cells in the most straightforward way — by fusing entire tumour cells with cultured dendritic cells.

The authors found that when cells from the murine MC38 carcinoma were fused with bone-marrow-derived dendritic cells, they obtained fusion hybrids. These were then used to immunize syngeneic (genetically identical) mice, and they generated a strong antitumour activity in these animals. Interestingly, this antitumour activity seemed

to involve both CD4- and CD8-positive T cells, in MHC class I- and II-restricted responses, respectively. This is in contrast to the largely CD8-positive, class I-restricted responses that have been observed when dendritic cells pulsed with antigen have been used as immunogens<sup>7</sup>. More excitingly — and potentially of greater clinical significance — this antitumour activity occurred not only against primary tumours, but also against pre-established metastatic disease<sup>3</sup>.

The fusion of tumour cells with antigen-presenting cells is not a new concept, and it has already been achieved with activated B cells<sup>9</sup>. However, as Kufe and colleagues point out, the use of dendritic cells offers two additional advantages — they can prime naive CTLs, and they are easily accessible from peripheral blood or bone marrow. This availability of donor cells would be of considerable benefit in any human application. But what of the tumour-cell side? Obviously, the single 'swallow' of the MC38 tumour does not a 'summer make', and both the procedure and the magnitude of the antitumour effects will need to be verified in further animal models. Moreover, efficacy against rodent tumours is not a guarantee of success against human cancers. The authors warn of the inherent limitations of animal models but, with their approach, it may be the technical rather than conceptual aspects that limit easy transfer to the clinic.

Fusion followed by selection means that, if the procedure is to be applied to individual human tumours, the neoplastic cells must have the capacity for continuous *in vitro* proliferation — not necessarily an attribute of all primary cell cultures. In the work by Kufe and colleagues, no selectable markers seem to have been incorporated into the fusion partners. This, of course, is not a problem when pure populations of tumour cells are used. But if mixed populations are derived from patient material, how can we guarantee that anti-self responses will not be generated against normal cells? These might fuse with the dendritic cells as easily as the cancer cells do, and lead to

undesirable anti-self responses on injection. One way to obviate such potential problems might be to create universal fuser lines of representative tumours, rather than relying on material from individual patients.

The work of Kufe and co-workers — with its dramatic effects on tumour burden and immunoprotection — is complementary to the encouraging results of other groups, which have used dendritic cells to stimulate naive T cells. For example, the pulsing of dendritic cells with tumour antigens *in vitro* generates both tumour-specific CTLs<sup>10</sup> and tumour resistance *in vivo*<sup>8</sup>. The rejection of established tumour metastases<sup>3,7–10</sup> parallels the anti-graft rejection responses in transplantation immunology induced by donor dendritic cells<sup>11</sup>. This potency of the immune response that is elicited by dendritic cells is also illustrated by their ability to break neonatal tolerance<sup>12</sup>.

These results are promising because they represent new approaches to targeted immunotherapy that are based on a greater appreciation of the mechanisms by which

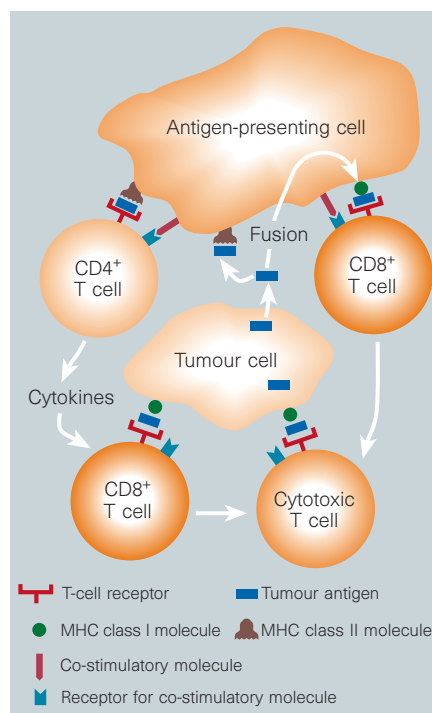


Figure 1 Kufe and colleagues<sup>2</sup> have developed a new strategy to activate the response of cytotoxic T lymphocytes against tumour cells. Pre-cytotoxic CD8-positive T lymphocytes may be primed directly by antigen-presenting cells, which present tumour antigens on a major histocompatibility complex (MHC) class I molecule. Presentation of tumour antigen by MHC class II molecules to CD4-positive helper T lymphocytes leads to the release of stimulatory cytokines, which activate the pre-cytotoxic CD8-positive T lymphocytes. By fusing one group of antigen-presenting cells — the dendritic cells — with entire tumour cells, the tumour antigens can be expressed and used to directly activate naive T lymphocytes.

tumour-specific cytolytic cells are generated. Whether this will eventually represent a turning point in the war against cancer, or whether it is yet another false dawn, considerably more work still needs to be done. But at least the initial battle-plans look hopeful, and there is reason for optimism. □

Ian Hart is in the Division of Oncology, United Medical Schools of Guy's and St Thomas' Hospitals, London SE1 7EH, UK. Camilo Colaco is at Quadrant Healthcare, Maris Lane, Cambridge CB2 2SY, UK.

## Biodiversity

# Global change through invasion

Gábor L. Lövei

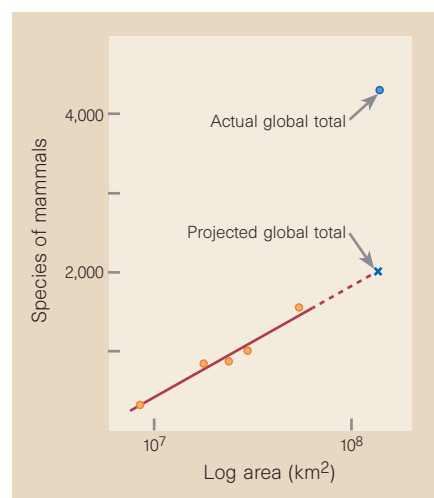
Invasion by plants and animals occurs everywhere, and has profoundly influenced the shape of the world's biota over geological time. From the very early stages of agriculture, humans have also intentionally transported crops and animals to other regions of the world, thus greatly advancing the material culture of different societies<sup>1</sup>; the introduction of many more organisms was not so welcome. Today, the scale of species introduction by humans is vastly increased. Vitousek *et al.*, in a paper in the *New Zealand Journal of Ecology*<sup>2</sup>, argue that species introductions should be recognized as an important component of human-induced global change and as a serious threat to biodiversity.

Invading plant and animal species have caused drastic changes in the receiving biota of islands like New Zealand or Hawaii, as testified by their recent history. Vitousek *et al.*'s carefully constructed<sup>3</sup>, extensive list of plant invasions now allows us to look at this phenomenon from a more global perspective. Introduced species make up 26–40 per cent of all plant species in isolated regions (Table 1), and even more on islands: for example, 82 per cent of Ascension Island's biota is composed of introduced species. Regions with a high volume of trade generally have more introduced species than less frequented areas. What is surprising, though, is that continental areas are no exceptions. Introduced species make a significant contribution both in terms of the number of species per unit

1. Bailer, J. C. & Gornik, H. L. *N. Engl. J. Med.* **336**, 1569–1574 (1997).
2. Gong, J., Chen, D., Kashiwaba, M. & Kufe, D. *Nature Med.* **3**, 558–561 (1997).
3. Vlasveld, L. T. & Rankin, E. M. *Cancer Treat. Rev.* **20**, 275–311 (1994).
4. Pardoll, D. M. *Curr. Opin. Immunol.* **5**, 705–706 (1993).
5. Marland, G. *et al. Stem Cells* **14**, 501–503 (1996).
6. Young, J. W. & Inaba, K. *J. Exp. Med.* **183**, 7–11 (1996).
7. Flamand, V. *et al. Eur. J. Immunol.* **24**, 605–610 (1994).
8. Boczkowski, D. *et al. J. Exp. Med.* **184**, 465–472 (1996).
9. Guo, Y. *et al. Science* **263**, 518–520 (1994).
10. Bakkar, A. B. H. *et al. Cancer Res.* **55**, 5330–5334 (1995).
11. Matzinger, P. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
12. Ridge, J. P. *et al. Science* **271**, 1723–1726 (1996).

area and as their relative share in the flora (Table 1). There are some striking examples for vertebrates too: New Zealand's fish fauna contains 30 'exotic' species (53 per cent of the total), Puerto Rico's has a whopping 91 per cent, and even Brazil has 13 per cent<sup>2</sup>.

Invaders frequently make a heavy impact. The eastern North American deciduous forests have suffered larger perturbations by invading pests and pathogens than from pollution or acid rain<sup>3</sup>. The long-term global effects of invasions are also significant, in two ways. First, they decrease the distinctiveness of local floras and faunas. New Zealand today has more vertebrate species than at the time of human colonization about 1,000 years ago<sup>4</sup>. Even if the extinctions are not considered (all but one of the 50 extinct species were endemic to New Zealand), only one of the introduced species, the Australian parma wallaby (*Macropus parma*), is threatened in its homeland<sup>5</sup>. So although New Zealand has become more 'diverse', it has also become more similar to the rest of the world. Second, geographic isolation is necessary for the maintenance of global biodiversity. For example, the area of a continent and the number of mammal species it supports show a tight correlation (Fig. 1). If all the Earth's dryland area formed one supercontinent, the predicted global total would be only about half of the actual global species richness (Fig. 1). Similar relationships have been found for several other groups of



**Figure 1** Species versus area curve for mammals. The number of species on a continent is tightly correlated with the size of the continent, but extrapolating that relation to the land area of Earth yields less than half the total number of species that actually occur on these continents. Much of the global diversity of mammalian species is due to the isolation of separate biotic regions. (Analysis prepared by A. Launer. Reproduced from ref. 2.)

organisms<sup>6</sup>. The current human-assisted massive invasions could, in fact, create one 'supercontinent', and with the geographical barriers broken down, the inevitable result would be a catastrophic loss of biodiversity.

Why have species introductions not been recognized as a global phenomenon? Acknowledged components of global change, like ozone depletion or carbon dioxide enrichment (the 'greenhouse effect'), occur on a very large spatial scale and simultaneously affect several different ecosystems. In contrast, invasions are mostly local, and involve a few species acting over a relatively short time frame. A more cohesive view was called for in order to demonstrate that although the players and the location of the ecological 'theatres' may be different, the same 'evolutionary play' is being continuously enacted all over the world. And we constantly busy ourselves with transporting the players from one theatre to another.

Possibly, because natural invasions are going on all the time, just like species extinctions, it was conceptually difficult to recognize the global implications. Species extinction is a problem not because it is a new phenomenon, but because its rate far exceeds the rate of speciation. Likewise, human-assisted invasions are effectively breaking down biogeographical barriers, and the resulting global mixing renders idle one of the most prolific means of generating biodiversity: isolation. If we cannot stop this McDonaldization of the biosphere, we stand to lose a substantial part of global biodiversity and may well find that we need variety not just as the spice of life — but for life itself. □

**Table 1** The contribution of non-native plant species to floras of different regions of the world

| Region/country          | Number of species |        | Exotic/(log) area | Percentage of exotics |
|-------------------------|-------------------|--------|-------------------|-----------------------|
|                         | Native            | Exotic |                   |                       |
| Russian Arctic          | 1,403             | 104    | 15.9              | 6.9                   |
| Europe                  | 11,820            | 721    | 102.8             | 5.7                   |
| United States           | 17,300            | 2,100  | 304.6             | 10.8                  |
| Southern Africa         | 20,573            | 824    | 128.1             | 3.9                   |
| Australia               | 15,638            | 1,952  | 283.5             | 11.1                  |
| Ontario                 | 2,056             | 805    | 133.5             | 28.1                  |
| New York state          | 1,940             | 1,083  | 210.7             | 35.8                  |
| Perth region, Australia | 1,510             | 547    | 136.1             | 26.6                  |
| British Isles           | 1,255             | 945    | 175.4             | 42.9                  |
| Hawaii                  | 1,143             | 891    | 210.9             | 43.8                  |
| New Zealand             | 2,449             | 1,623  | 298.9             | 39.9                  |

Modified from Table 1 in Vitousek *et al.*<sup>2</sup>.



Gábor L. Lövei is in the Horticulture and Food Research Institute, Private Bag 11030, Palmerston North, New Zealand.

1. Diamond, J. M. *Guns, Germs, and Steel: The Fate of Human Societies* (Norton, New York, 1997).
2. Vitousek, P. M., D'Antonio, C. M., Loope, L. D., Rejmánek, M. & Westbrooks, R. *NZ J. Ecol.* **21**, 1–16 (1997).

3. Campbell, F. T. & Schlarbaum, S. E. *Fading Forests: North American Trees and the Threat of Exotic Pests* (Nat. Res. Defense Council, New York, 1994).
4. Wilson, K.-J. *Pacific Cons. Biol.* (in the press).
5. King, C. M. (ed.) *Handbook of New Zealand Mammals* (Oxford Univ. Press, 1990).
6. Wright, D. H. *Int. J. Biometeorol.* **331**, 293–299 (1987).

## Acoustics

# Unwanted and wanted sound

Thomas D. Rossing

More and more people are becoming concerned about the control of environmental noise. That was certainly true of those attending the joint meeting of the Acoustical Society of America and the Institute of Noise Control Engineering in June\*. Although the meeting included more than 100 sessions covering all aspects of acoustics, noise control received particular attention — especially active control of noise and vibration.

In active noise control, a controlled source is used to cancel out an unwanted sound or vibration. It may employ either feedback or feedforward control (C. Hansen, Adelaide Univ.). In a duct, for example, a basic feedback system uses a microphone and a loudspeaker, along with an electronic controller. The loudspeaker vibrates to cancel out as much of the source sound as possible and to produce a minimum sound level at the error microphone (Fig. 1a). In a basic feedforward system, a reference microphone is added to sample the source noise before it reaches the error microphone (Fig. 1b). Because of their inherent stability, feedforward controllers are generally preferred over feedback controllers. Adaptive logic and other refinements can be added.

Effective noise control often requires the reduction of noise transmission through panels or partitions. In active structural acoustic control, actuators are attached to a panel to change its radiation characteristics. For example, 'smart' foam, with a piezoelectric actuator in between layers of sound-absorbing foam, can reduce sound radiation from vibrating plates (C. Guigou and C. R. Fuller, Virginia Polytechnic Inst.). And a new active interior trim panel can control aircraft noise at low frequency, up to 400 Hz (S. Sharp, G. Koopmann and W. Chen, Penn State Univ.). The actuator is a small, rigid panel that is driven in a piston-like motion by a moving-coil transducer similar to that of a loudspeaker.

In contrast to these papers on unwanted sound, a special session on loudspeakers dealt with improving the quality of wanted sound. Gabriel Weinreich (Univ. Michigan) has designed and built a loudspeaker with a direc-

tional radiation pattern that is a strong function of angle and frequency, like a violin's output at frequencies above 850 Hz. This produces the illusion that each note comes from a slightly different direction, confusing the common psychoacoustic cues and endowing the sound with a striking spatial sense like that of a violin, an instrument especially noted for its 'directional tone colour'. Weinreich demonstrated the ability of this loudspeaker to reproduce the sound of a violin and a pipe organ. Conversely, an ear-catching demonstration by Bruce Edgar, using his own voice, showed how a new horn design could be used to remove the sound colorations that plagued many vintage horn systems.

Sound can be 'wanted' for rather different purposes. By recording an incoming sound wave, reversing its time structure and re-emitting it, it can be focused back to its point of origin (this time-reversal technique has precedents in radar and optical technology). Practical time-reversal mirrors consist of large, ultrasonic transducer arrays. They have applications in medicine (to destroy kidney stones, for example) and in non-destructive testing of solids (M. Fink, Univ. Denis Diderot, Paris).

Time-reversal mirrors also have applications in the sea, but here the method is complicated by the fact that the medium is changing with time. Yet two experiments conducted in the Mediterranean Sea in April

1996 and May 1997 demonstrated that an acoustic time-reversal mirror can focus sound back to its origin even in an inhomogeneous ocean. William Kuperman (Scripps Instn, San Diego) and colleagues used 20 sources and 20 receivers in a vertical array 77 m across. A source that sent out a 50-ms ping every two minutes was towed through the focus of the array. The signal was received, reversed and retransmitted successfully to distances as great as 30 km. This experimental result is particularly remarkable when one considers that a spot 1/10th as deep as the waveguide is projected to a range more than 300 times the depth, with a horizontal extent of about one waveguide depth. These methods can be used to learn more about acoustic propagation in the ocean, perhaps for communications purposes, and to study fluctuations in currents, temperature and salinity using sound as a probe.

Returning to airborne sound of the desirable variety, Judith Brown (Wellesley College and MIT) discussed how to use computers to identify the sounds of musical instruments. The most successful schemes for identifying human speakers rely on the formant structure of speech sounds. (Formants are resonances of the vocal tract that determine speech sounds. Each vowel sound, for example, is characterized by three to five formant frequencies.) Identifying the formants of each speaker allows a computer to distinguish speakers through pattern recognition.

So Brown followed the same procedure for the oboe and saxophone. It turns out that a computer can do very well — in fact, it scored considerably better than humans in identifying oboe sounds, although in the case of saxophones there was little difference. So, as well as doing clever things with sound production, some automatic systems can appreciate the result for us. □

Thomas D. Rossing is in the Department of Physics, Northern Illinois University, DeKalb, Illinois 60115-2854, USA.

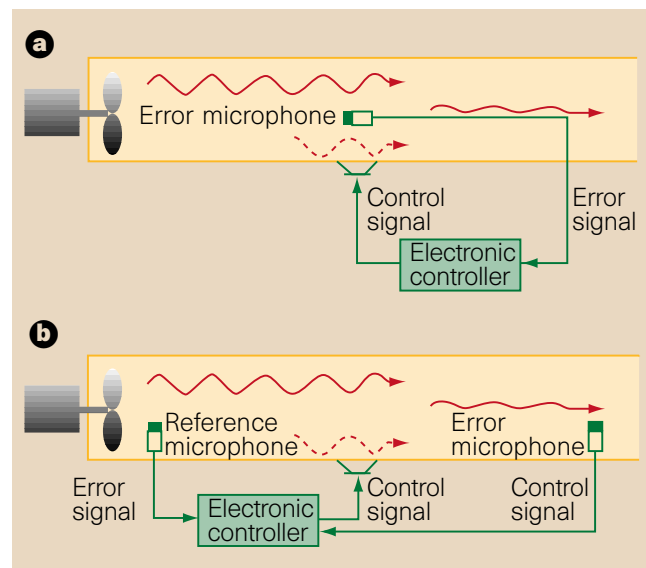


Figure 1 Two schemes for active noise control. a, Simple feedback, in which the control signal is tweaked to cancel out the unwanted noise; b, 'feedforward', in which a second microphone samples the noise, giving a rough value for the control signal directly, and so avoiding the instabilities of simple feedback.

\* 133rd Meeting of the Acoustical Society of America, held jointly with NOISE-CON 97 (National Conference on Noise Control Engineering), Pennsylvania State University, State College, Pennsylvania, USA, 14–20 June 1997.

## LDL-receptor structure

# Calcium cages, acid baths and recycling receptors

Michael S. Brown, Joachim Herz and Joseph L. Goldstein

How would you design a 40-amino-acid peptide that is rigid enough to survive hundreds of acid baths at 10-minute intervals, yet adaptable enough to bind unrelated macromolecules with high affinity at pH 7.4, and discharge them rapidly at pH 5.5? If you answered "Put in three disulphide bonds and an octahedral calcium cage", you were right, but too late — nature invented this peptide more than 500 million years ago, and has used it repeatedly ever since. The peptide in question is the ligand-binding repeat of the low-density lipoprotein receptor (LDLR), and the basis for its stability is revealed in elegant detail by the crystallographic studies of Fass *et al.*<sup>1</sup> on page 691 of this issue.

The LDLR is the patriarch of a family of cell-surface receptors (Fig. 1) that transport macromolecules into cells by receptor-mediated endocytosis in clathrin-coated pits (specialized intucked regions of the plasma membrane)<sup>2</sup>. Every 10 minutes, the LDLR completes a cycle in which it binds a ligand at the cell surface, moves to an acidic endosome where it discharges its ligand, then returns to the surface. The receptor's ligand-binding domain comprises seven imperfect repeats of about 40 amino acids. Each repeat contains six cysteine residues which are disulphide bonded in the pattern one to three, two to five, and four to six<sup>3</sup>. Reduction of these disulphides destroys the structure and abolishes binding<sup>2</sup>.

The disulphide bridges are not the only reason for the stability of the LDLR. Ligand binding is destroyed by metal chelators and restored by calcium<sup>4</sup>. Fass *et al.*<sup>1</sup> explain this by showing that each repeat contains a single  $\text{Ca}^{2+}$  ion trapped in an octahedral cage that is formed by four conserved acidic residues plus two nearby carbonyl oxygens. Evidence that this  $\text{Ca}^{2+}$  cage stabilizes the LDLR is consistent with previous studies<sup>5</sup> of patients with familial hypercholesterolaemia, who have mutations that destroy LDLR function. The LDLR normally removes cholesterol-carrying lipoproteins from blood, so these patients have high levels of blood cholesterol, and suffer from heart attacks at a young age. More than 250 LDLR mutations have been described, among which about 50 are point mutations that substitute individual amino acids in the seven repeats of the ligand-binding domain<sup>5</sup>. Many of the point mutations replace the residues of the  $\text{Ca}^{2+}$  cage, thereby destabilizing the receptor.

The LDLR binds two ligands — apolipoprotein E (apoE;  $M_r$  34,000) and apolipoprotein B (apoB;  $M_r$  513,000). Although both ligands are found on the

surface of plasma lipoproteins, they differ in amino-acid sequence and have no known structural relationship other than their common affinity for negatively charged polymers such as glycosaminoglycans. LDL contains one copy of apoB, but other lipoproteins that bind to the LDLR contain many apoEs. ApoE-binding primarily requires repeat 5 of the LDLR, the inactivation of which reduces binding by 60 per cent<sup>6</sup>. ApoB-binding, on the other hand, requires a combination of repeats 3 to 7, plus adjacent cysteine-rich repeats from a different family. So different combinations of repeats allow a single receptor to recognize structurally diverse ligands.

Crystallographic studies have identified the binding site on apoE to be a positively charged, amphipathic helix<sup>7</sup>. When the positive charges on this helix were abolished by mutation, reductive methylation or cyclohexanedione modification, binding of apoE

to the LDLR was abolished<sup>8</sup>. Similar results were obtained by chemical treatment of apoB in LDL, and binding of LDL to purified LDLR was prevented by polyanions such as suramin<sup>9</sup>. So the binding reaction seems to be based on ionic interactions between positively charged helices on the apolipoprotein ligands, and negative residues in the ligand-binding repeats. But these binding studies raise a paradox when considered in the light of the structure provided by Fass *et al.*<sup>1</sup>.

The conundrum is that, by crystallographic analysis of repeat 5 in the LDLR, Fass *et al.* reveal that most of the negatively charged residues are coordinated with  $\text{Ca}^{2+}$  and are, therefore, unavailable to bind to apoE or apoB. Perhaps the apolipoproteins interact with other negatively charged residues such as Asp32 (see Fig. 1 on page 691), which is highly conserved in the repeats<sup>6</sup>. It is also possible that apoE and apoB bind to the interfaces between the repeats, or that the structure of a single repeat becomes altered when it is present in a chain with other repeats. For definitive answers, we need to know the structure of the ligand-receptor complex.

As well as the ligand-binding repeats, the LDLR contains a 411-amino-acid sequence

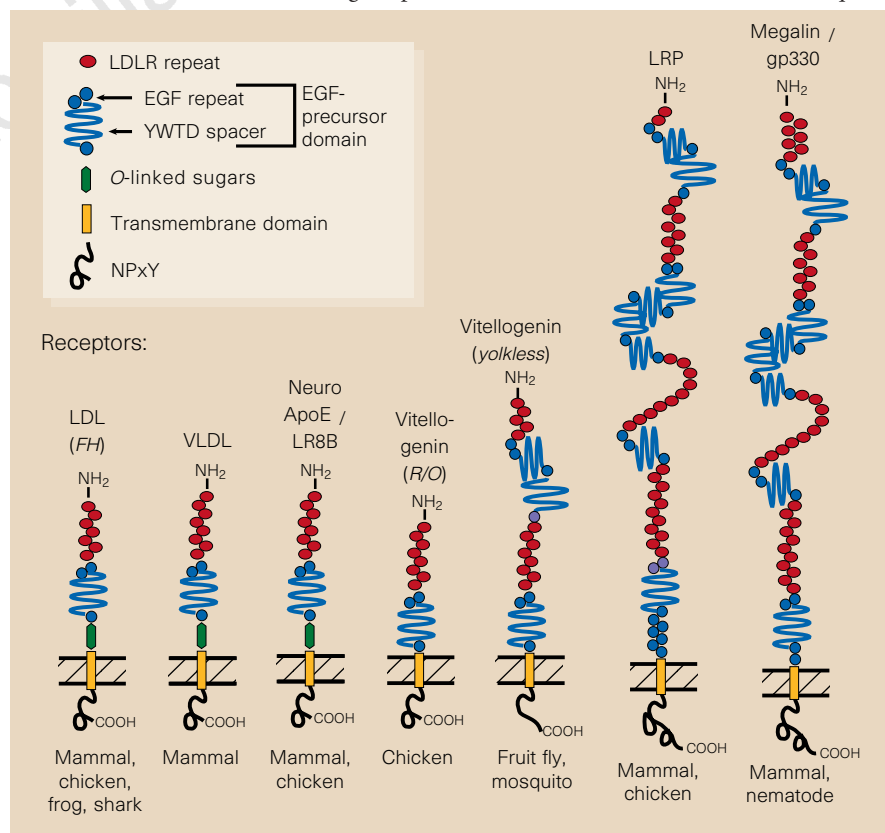


Figure 1 Members of the low-density lipoprotein receptor (LDLR) gene family share extracellular sequences (ligand-binding repeats and epidermal growth factor (EGF)-precursor domains, with or without O-linked sugar domains), transmembrane orientation, and cytoplasmic NPXY sequences that mediate the internalization of coated pits. Fass *et al.*<sup>1</sup> have solved the crystal structure of ligand-binding repeat 5 of the LDLR, and they find that the basis for its stability is an octahedral cage containing  $\text{Ca}^{2+}$ . Naturally occurring mutations that disrupt receptor function are denoted in parentheses: FH, familial hypercholesterolaemia in humans and rabbits; R/O, restricted ovulator in chickens; and yolkless in *Drosophila melanogaster*.

that is shared with the membrane-bound precursor of epidermal growth factor (EGF). The domain consists of two, 40–50-amino-acid, cysteine-rich repeats of the EGF class, then five copies of a roughly 50-amino-acid (cysteine-poor) repeat which is characterized by the consensus sequence YWTD. This is followed by a third, cysteine-rich, EGF repeat (Fig. 1). When the EGF-precursor domain is deleted from the LDLR, the receptor can no longer bind LDL but it still binds lipoproteins that contain apoE<sup>10</sup>. The bound ligands enter the cell as normal, but they are not released in the endosome — instead, the receptor is degraded along with the ligand. *In vitro* binding studies have shown that the deleted receptor fails to release its ligand at acid pH, thereby explaining its failure to recycle<sup>10</sup>.

Although the EGF repeats each contain three disulphide bonds, they otherwise differ from the ligand-binding repeats. For example, an NMR structure of two adjacent EGF repeats from fibrillin reveals that, rather than being trapped in a cage, Ca<sup>2+</sup> bridges the gap between repeats<sup>11</sup>. The mechanism by which the EGF-precursor region of the LDLR facilitates dissociation of lipoproteins from the ligand-binding region is not known. But, at acid pH, the EGF region probably undergoes a conformational change that is mediated either by the dissociation of Ca<sup>2+</sup> or by titration of the aspartic acids in the YWTD repeats.

The third important domain of the LDLR is the 50-amino-acid cytoplasmic tail, which contains the sequence NPxY (where x is any amino acid; Fig. 1). This sequence forms a reverse turn<sup>12</sup>, and it clusters the receptor into coated pits, apparently by interacting with a clathrin-associated protein<sup>13</sup>. Phosphorylation of the tyrosine residue in NPxY is not required for coated-pit clustering — phenylalanine can substitute without effect<sup>13</sup>.

There are, to date, seven members of the LDLR family (Fig. 1). Although other proteins contain LDLR ligand-binding repeats or EGF repeats, these seven are the only proteins that contain both. All except the insect varieties also share the NPxY signature in the cytoplasmic tail. In the insect receptors, NPxY is replaced by a dileucine motif<sup>14</sup>, which can also mediate coated-pit clustering.

The behemoths of the family are the LDLR-related protein (LRP) and megalin/gp330. Megalin (4,660 amino acids) contains 36 ligand-binding repeats and eight EGF-precursor domains, the arrangement of which is almost completely conserved from nematodes to mammals<sup>15</sup>. Each member of the LDLR family binds many ligands, and the number of ligands bound seems to increase with the number of ligand-binding repeats. The champion is LRP, which has at least 20 known ligands ranging from lipoproteins (chylomicron remnants) and protease inhibitors ( $\alpha_2$ -macroglobulin) to plasminogen activators<sup>15</sup>. Each macromolecule probably binds to a unique combination of

ligand-binding repeats.

Members of the LDLR family can transfer large amounts of macromolecules into cells due to the combination of Ca<sup>2+</sup>-stabilized binding repeats, acid-regulated EGF-precursor regions and cytoplasmic internalization signals. For example, the vitellogenin receptor transports 1.5 g of protein into the chicken oocyte each day, as the single cell grows to the size of a grade AA jumbo egg over seven days. A mutation in this receptor causes the female-sterile 'restricted ovulator' phenotype, in which chicken oocytes do not grow to normal size<sup>16</sup>. A mutation in the *Drosophila* homologue of the vitellogenin receptor also produces a female-sterile phenotype, *yolkless*<sup>14</sup>.

Now that Fass *et al.*<sup>1</sup> have solved the structure of a single ligand-binding repeat, we need to know how the repeats are orientated with respect to each other and to the other domains. In fibrillin, where Ca<sup>2+</sup> bridges two adjacent EGF repeats, the repeats are probably stacked in an extended fashion, perhaps in the form of a helix<sup>11</sup>. And as the ligand-binding repeats of the LDLR are separated by short, 4–5-amino-acid sequences which do not allow for much flexibility, these repeats could also be stacked atop one another. The exception is a 12-residue spacer between repeats 4 and 5 which, in LDLRs from all known species, is conserved in length but not in sequence<sup>17</sup>. Does this spacer form a loop that allows the stacks of repeats to double back on one another? The answer will surely be forthcoming now that Fass *et al.* have taught us how to assemble, crystallize and appreciate a rigid, disulphide-bonded Ca<sup>2+</sup> cage. □

Michael S. Brown, Joachim Herz and Joseph L. Goldstein are in the Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA.

1. Fass, D., Blacklow, S., Kim, P. S. & Berger, J. M. *Nature* **388**, 691–693 (1997).
2. Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W. & Schneider, W. J. *Annu. Rev. Cell Biol.* **1**, 1–39 (1985).
3. Bieri, S., Djordjevic, J. T., Daly, N. L., Smith, R. & Kroon, P. A. *Biochemistry* **34**, 13059–13065 (1995).
4. Goldstein, J. L. & Brown, M. S. *J. Biol. Chem.* **249**, 5153–5162 (1974).
5. Goldstein, J. L., Hobbs, H. H. & Brown, M. S. in *The Metabolic and Molecular Bases of Inherited Disease* 7th edn (eds Scriver, C. R. *et al.*) 1981–2030 (McGraw-Hill, New York, 1995).
6. Russell, D. W., Brown, M. S. & Goldstein, J. L. *J. Biol. Chem.* **264**, 21682–21688 (1989).
7. Wilson, C., Wardell, M. R., Weisgraber, K. H., Mahley, R. W. & Agard, D. A. *Science* **252**, 1817–1822 (1991).
8. Mahley, R. W. & Rall, S. C. Jr in *The Metabolic and Molecular Bases of Inherited Disease* 7th edn (eds Scriver, C. R. *et al.*) 1953–1980 (McGraw-Hill, New York, 1995).
9. Schneider, W. J., Beisiegel, U., Goldstein, J. L. & Brown, M. S. *J. Biol. Chem.* **257**, 2664–2673 (1982).
10. Davis, C. G. *et al.* *Nature* **326**, 760–765 (1987).
11. Downing, A. K. *et al.* *Cell* **85**, 597–605 (1996).
12. Bansal, A. & Gierasch, L. M. *Cell* **67**, 1195–1201 (1991).
13. Chen, W.-J., Goldstein, J. L. & Brown, M. S. *J. Biol. Chem.* **265**, 3116–3123 (1990).
14. Schonbaum, C. P., Lee, S. & Mahowald, A. P. *Proc. Natl. Acad. Sci. USA* **92**, 1485–1489 (1995).
15. Krieger, M. & Herz, J. *Annu. Rev. Biochem.* **63**, 601–637 (1994).
16. Schneider, W. J. *Int. Rev. Cytol.* **166**, 103–137 (1996).
17. Mehta, K. D., Chen, W., Goldstein, J. L. & Brown, M. S. *J. Biol. Chem.* **266**, 10406–10414 (1991).

## Daedalus

### Command performance

Some while ago Daedalus proposed the use of pulsed magnetic fields in psychotherapy. Unlike electro-shock therapy, they induce purely local currents in the brain. The technique now seems to be maturing. It is claimed that pulsed magnetic fields, applied to specific regions of the brain, can affect motor control or relieve depression. So Daedalus is now sharpening the selectivity of the method, and extending it to the nervous system as a whole.

A tightly shaped magnetic pulse will induce current in, and therefore fire, nerves in quite a small region of tissue. By itself, this might not be useful; even a narrow nerve trunk can carry thousands of individual fibres, each with a different destination. But Daedalus feels that each fibre must have its own 'resonant frequency', possibly quite broad, at which it will fire most effectively. It may also respond best to a specific shape of pulse. So if magnetic pulses of carefully adjusted frequency, amplitude and shape are aimed at a nerve trunk, they should trigger just one class of fibre (those to a specific muscle, say) without affecting the rest.

Magnetic pulsing could wonderfully improve muscular coordination and strength. Daedalus recalls an electrician who claims the world record for throwing the pliers. He touched a high-voltage terminal with them, and every muscle-fibre in his arm contracted at once, more forcefully than he could have managed by any act of will. So Daedalus is devising a magnetic-pulse machine to be placed round an athlete's neck. Without the pain and imprecision of electric stimulation, it will take over the spinal nerves leading to his arms and legs. A well-designed 'neuromagnetic program', firing the right nerves in the right sequence, would then let him run, jump or throw with ultimate power and accuracy. An Olympic committee would probably disallow these achievements. Even so, by training with the system, he could hone his own voluntary action to utter perfection.

Sensory nerves could also be triggered, or indeed inhibited. At present, implanted coils are used to block pain, for example from damaged joints. A magnetic inhibitor would be far less invasive, and could be turned on or adjusted in seconds. It could even abolish pain while leaving sensation untouched. As centralized medical services run down, it could become an ideal surgical aid. With a magnetic inhibitor, a scalpel, a mirror and a DIY manual, you could remove your own appendix.

David Jones



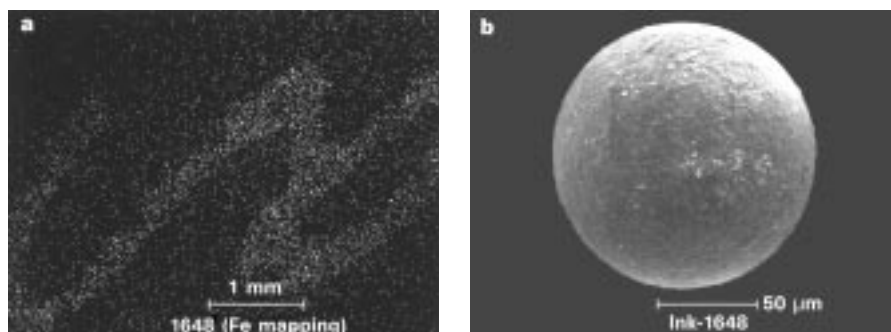
# Framboidal pyrites in antique books

During the Middle Ages and the Renaissance, ink was commonly manufactured by mixing tannin with iron sulphates. The anoxic environment inside ancient books favours the reduction of the sulphate in the ink and allows spherical aggregates (framboids) of submicrometre-sized pyrite crystals (iron sulphide) to be formed.

Framboidal pyrites are ubiquitous in a variety of conditions and geological environments such as hydrothermal veins and sedimentary rocks<sup>1,2</sup>. The meaning of the term 'framboid' has changed significantly since the beginning of the century, when it was used to emphasize a bacteriogenic origin. Since then, non-organic formation of framboids has been shown experimentally. The term 'framboid' is now used to describe spherical aggregates of microcrystals with a diameter of up to 150  $\mu\text{m}$  (ref. 3).

For the formation of framboids, the precipitation of iron and sulphur ions in a reducing environment is required. We have found that such an environment exists in books from the sixteenth and seventeenth centuries that have been stored in archives for hundreds of years. We discovered these authigenic framboidal pyrites in books during a restoration programme at the Archivo Histórico Nacional in Madrid. We found a shiny black powder in the seams joining the pages to the spines of the books. Studying this powder, by transmitted and reflected light microscopy, X-ray diffraction and scanning electron microscopy, revealed it to be solid remains of the ink. This was corroborated by the presence of the same remains adhering to the writing itself (Fig. 1a).

The mineral phases included quartz, haematite, K-feldspar, allochthon pyrites, calcite, gypsum, rutile, magnetite and ilmenite, with crystal sizes ranging from 50 to 400  $\mu\text{m}$ ; additives commonly used to improve the finish of inks at that time<sup>4</sup>. Authigenic pyrite was found in this black powder as framboidal aggregates of  $\sim 140 \mu\text{m}$  in diameter, made up from tiny, subhedral, pyrite microcrystals. These framboids (Fig. 1b) are nearly perfect spheroids with smooth, homogeneous surfaces, caused by the coales-



**Figure 1** Scanning electron micrographs **a**, Iron mapping (SEM-EDAX) which clearly follows the shape of a letter in the text. **b**, Framboidal pyrites formed within the book.

cence of the pyrite microcrystals, indicating a certain degree of framboidal evolution<sup>2</sup>.

It has recently been stated<sup>5</sup> that: framboid size is controlled by the residence time near the oxic–anoxic boundary; growth times for framboids in the water columns of euxinic basins are less than three months; and diagenetic framboids grow for about three times longer than syngenetic framboids. The large size of our framboids (140  $\mu\text{m}$ ) indicates that optimum growth conditions occur in these old books. The presence of red rot on the cover and some of the pages shows that there are euxinic micro-environments inside, and oxic–anoxic boundaries could have been maintained by the seasonal changes in Madrid (temperatures from  $-10^{\circ}\text{C}$  to  $44^{\circ}\text{C}$ ). The period inside the books was long considering the calculated framboid growth rate ( $4 \mu\text{m yr}^{-1}$ ) in euxinic environments<sup>5</sup>.

The chemical elements of these pyrite framboids almost certainly came from the components of the ink: tannin, gum arabic, iron salts, additives and solvents. Iron gallic inks used at the time were made by the reaction of tannins, derived from gallic acid ( $\text{C}_6\text{H}_2(\text{OH})_3\text{COOH}$ ) with iron sulphates (melanterite  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ )<sup>6</sup>. Cellulose in the paper and gum arabic used in the ink provided carbon and nitrogen needed to develop iron- and sulphate-reducing bacteria. These, in turn, contributed to the formation of framboidal pyrites: framboidal pyrite has been found in the similar cellulose-rich environment of several species of fossil woods<sup>7</sup>.

As organic matter is present in the closed environment of the books, the possibility of sulphate- and iron-reducing bacteria (such as *Desulfovibrio desulfuricans*, *Desulfotomaculum nigrificans*) being responsible for pyrite mineralization should also be considered<sup>8</sup>.

The detection of processes of pyrite mineralization in this singular environment shows that unusually large pyrite framboids, displaying a high degree of evolution, can be formed under intermittent oxic and anoxic conditions over hundreds of years. This discovery supports the recently published hypothesis that the formation of framboidal pyrites does not require a narrow set of physical or chemical conditions<sup>1</sup>.

**Javier García-Guinea, Jesus Martínez-Frias  
Rafael Gonzalez-Martin**

Museo Nacional de Ciencias Naturales CSIC,  
C/ Jose Gutierrez Abascal 2, 28006-Madrid, Spain  
e-mail: guinea@fresno.csic.es

**Luis Zamora**

Archivo Histórico Nacional CSIC,  
C/ Serrano 115, 28006-Madrid, Spain

1. Wilkin, R. T. & Barnes, H. L. *Geochim. Cosmochim. Acta* **61**, 323–339 (1997).
2. Martínez-Frias J., Navarro-Flores, A., Lunar-Hernández, R. *Neues Jb. Miner. Mh.* **4**, 175–184 (1997).
3. Ferrand, M. *Miner. Deposita* **5**, 237–247 (1970).
4. Zerdoun Bat-Yehouda, M. (ed.) *Les Encre Noires au Moyen Age* (CNRS, Paris, 1983).
5. Wilkin, R. T., Barnes, H. L. & Brantley, S. L. *Geochim. Cosmochim. Acta* **60**, 3897–3912 (1996).
6. Wunderlich, C. H., Weber, R. & Bergerhoff, G. *Z. Anorg. Allg. Chem.* **598**, 371–376 (1991).
7. García-Giménez, R. et al. *Bol. Soc. Esp. Miner.* **7**, 115–122 (1983).
8. Mann, S. et al. *Nature* **343**, 258–261 (1990).

## Intercropping increases parasitism of pests

As part of a programme for controlling lepidopteran stem-borers in cereal crops in Africa, we have investigated the effectiveness of combined cropping regimes of cultivated and wild plants for reducing stem-borer damage. Intercropping with the

non-host molasses grass, *Melinis minutiflora*, significantly decreased levels of infestation by stem-borers in the main crop and also increased larval parasitism of stem-borers by *Cotesia sesamiae*. Volatile agents produced by *M. minutiflora* repelled female stem-borers and attracted foraging female *C. sesamiae*. One of the volatile components released by intact *M. minutiflora* which attract parasitoids is also produced by herbivore-damaged plants and is impli-

cated more widely as a cue for stimulating predation and parasitism.

Maize (*Zea mays*) and sorghum (*Sorghum bicolor*) are the most important cereal crops for the people of Africa. Lepidopteran stem-borers are ubiquitous pests that attack these crops throughout their growth stages and the larvae cause damage ranging from 20 to 80% loss of yield. One approach to pest control in resource-poor regions is to develop management systems

**Table 1 Response of *C. sesamiae* to plant or plant extract**

| Stimulus            | Response  |         | Significance ( <i>P</i> ) |
|---------------------|-----------|---------|---------------------------|
|                     | Treatment | Control |                           |
| Live plant (30 g)   | 39        | 7       | <0.005                    |
| Extract (0.2 mg)    | 22        | 18      | Not significant           |
| Extract (2 mg)      | 32        | 18      | <0.05                     |
| Extract (20 mg)     | 44        | 9       | <0.005                    |
| Nonatriene (0.1 mg) | 35        | 18      | <0.05                     |
| Nonatriene (1 mg)   | 36        | 14      | <0.005                    |

Response of female *Cotesia sesamiae* in the Y-tube olfactometer to *Melinis minutiflora* (live plant or extract) and (*E*)-4,8-dimethyl-1,3,7-nonatriene. Number of females choosing each 'arm' of the Y-tube is given. Control was clean air. 1 mg plant extract=20.1g (wet mass) plant material.

using the 'push-pull' or stimuldeterrent diversionary strategy<sup>1</sup>, whereby insects are repelled from a harvestable crop and simultaneously attracted to a 'discard' or 'trap' crop. For maximum efficacy, these systems should also exploit natural enemies, particularly hymenopteran parasitoids, which can be important in suppressing pest populations<sup>2</sup>. Indeed, reductions in such beneficial organisms frequently trigger pest outbreaks<sup>3</sup>.

To develop a diversionary strategy for small-scale African cereal production, we assessed a range of cultivated and wild plants in the Gramineae family (Poaceae) in field trials in Kenya for susceptibility to stem-borers, particularly the indigenous *Busseola fusca* (Lepidoptera, Noctuidae) and the introduced *Chilo partellus* (Lepidoptera, Pyralidae). In these trials, molasses grass showed no colonization by stem-borers. Further, volatiles extracted by hydro-distillation of the plant repelled gravid female stem-borers in a laboratory oviposition assay (for *C. partellus*, eggs laid per filter-paper disc: control 40.9, 100 µg *M. minutiflora* extract 2.2;  $P < 0.005$ ,  $n = 8$ ). In field trials at Mbita Point on Lake Victoria, *M. minutiflora* planted in alternate rows with maize significantly reduced stem-borer infestation of the main crop (damaged maize plants: single crop 39.2%, intercropped with *M. minutiflora* 4.6%;  $P < 0.01$ ). We also saw a significant increase in parasitism by the larval parasitoid *C. sesamiae* (Hymenoptera, Braconidae) (parasitized larvae in maize: single crop 5.4%, maize with *M. minutiflora* intercrop 20.7%;  $P < 0.01$ ).

To identify the chemicals mediating this behaviour of stem-borers and parasitoids, we isolated volatiles from live *M. minutiflora* plants by entrainment into porous polymer<sup>4</sup>. Electrophysiologically active components in the solvent-eluted samples were located by coupled gas chromatography and electroantennography<sup>5</sup>. We tentatively identified active peaks by gas chromatography-mass spectrometry and confirmed their identity by co-injection with authentic compounds on two columns of different polarity, and using behavioural studies<sup>6</sup>. Characterized semiochemicals included α-terpinolene, the ocimene isomers, β-caryophyllene, humulene and (*E*)-4,8-dimethyl-1,3,7-nonatriene.

Production of some of the compounds released by intact *M. minutiflora* can also be

induced in plants damaged by herbivorous insects<sup>7-9</sup>. The nonatriene in particular has been implicated as an 'SOS' signal recruiting predators and parasites<sup>10</sup>. The presence of such compounds in the *M. minutiflora* intercropping system could provide an explanation for the increased parasitism observed. In behavioural assays using a Y-tube olfactometer<sup>6</sup>, we showed that foraging female *C. sesamiae* were indeed attracted to live *M. minutiflora* plants and also responded in a dose-dependent manner to the hydrodistillation extract (Table 1), and to the nonatriene alone.

The prospects for understanding and exploiting the interaction of hymenopteran parasitoids with their hosts have advanced rapidly, particularly with the discovery that semiochemicals released during herbivore damage can stimulate parasitoid foraging<sup>11-13</sup>. Our study suggests that intact plants with an inherent ability to release such stimuli could be used in new crop protection strategies.

**Z. R. Khan, K. Ampom-Nyarko  
P. Chiliswa, A. Hassanali, S. Kimani  
W. Lwande, W. A. Overholt**

*International Centre of Insect Physiology  
and Ecology,  
PO Box 30772, Nairobi, Kenya*

**J. A. Pickett\*, L. E. Smart  
L. J. Wadhams, C. M. Woodcock**  
*IACR-Rothamsted, Harpenden,  
Hertfordshire AL5 2JQ, UK  
e-mail: John.Pickett@bbsrc.ac.uk*

1. Miller, J. R. & Cowles, R. S. *J. Chem. Ecol.* **16**, 3197-3212 (1990).
2. Pickett, J. A., Wadhams L. J. & Woodcock, C. M. *Agriculture, Ecosystems and the Environment* (Elsevier, in the press).
3. Debach, P. & Rosen, D. (eds) *Biological Control by Natural Enemies* (Cambridge Univ. Press, 1991).
4. Blight, M. M. in *Chromatography and Isolation of Insect Hormones and Pheromones* (eds McCaffery, A. R. & Wilson, I. D.) 289-298 (Plenum, New York, 1990).
5. Wadhams, L. J. in *Chromatography and Isolation of Insect Hormones and Pheromones* (eds McCaffery, A. R. & Wilson, I. D.) 281-288 (Plenum, New York, 1990).
6. Ngi-Song, A. J. et al. *J. Chem. Ecol.* **22**, 307-323 (1996).
7. Takabayashi, J., Dicke, M. & Posthumus, M. A. *J. Chem. Ecol.* **20**, 1329-1354 (1994).
8. Potting, R. P. J., Vet, L. M. & Dicke, M. *J. Chem. Ecol.* **21**, 525-539 (1995).
9. Agelopoulos, N. G. & Keller, M. A. *J. Chem. Ecol.* **20**, 1955-1967 (1994).
10. Dicke, M. *J. Plant Physiol.* **143**, 465-472 (1994).
11. Lewis, W. J. & Tumlinson, J. H. *Nature* **331**, 257-259 (1988).
12. Turlings, T. C. J., Tumlinson, J. H. & Lewis, W. J. *Science* **250**, 1251-1253 (1990).
13. Turlings, T. C. J. & Tumlinson, J. H. *Proc. Natl Acad. Sci. USA* **89**, 8399-8402 (1992).

\*To whom correspondence should be addressed.

## Stochastic resonance at the single-cell level

Does the electricity with which we power our world pose significant health hazards from the resulting electromagnetic fields? Several authors<sup>1,2</sup> have speculated that 'stochastic resonance'<sup>3</sup> — a nonlinear phenomenon in which the addition of noise to a system increases its response to an external signal — may allow biological cells to detect and respond to very weak external electric fields far below the thermal noise limit<sup>4,5</sup>, and thus possibly cause harmful effects. Here we examine this question using a recent theory of Bezrukov and Vodyanoy<sup>6</sup> for the effect of non-equilibrium noise on a voltage detector (such as a biological ion channel<sup>7</sup>). We show that with parameters appropriate for typical biological cells, adding noise does not make a far-from-detectable signal detectable.

Bezrukov and Vodyanoy imagine a Poisson process (for example, the entry of a toxic molecule through a protein gate<sup>8</sup>), the rate of which is modulated by an external low-frequency signal voltage with dimensionless amplitude  $V_s$  and a zero average gaussian noise voltage  $V_N(t)$ . For a small-amplitude signal, the signal-to-noise ratio (SNR) is:

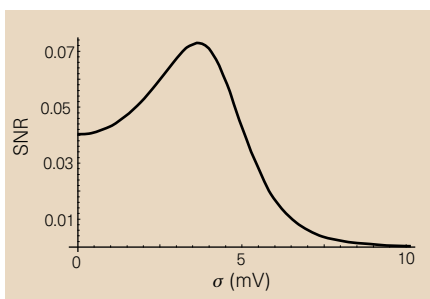
$$\text{SNR} = \frac{V_s^2}{2\Delta f_a} r(0) \exp\left(\frac{\sigma^2}{2}\right) \quad (1)$$

$$2 + \frac{r(0)}{f_c} \exp\left(\frac{\sigma^2}{2}\right) \sum_{n=1}^{\infty} \frac{\sigma^{2n}}{n!n}$$

where  $r(0)$  is the basal rate of the Poisson process,  $\Delta f_a$  is the bandwidth of the detector, and  $f_c$  is the corner frequency and  $\sigma$  the dimensionless r.m.s. amplitude of the external noise.

The SNR is a non-monotonic function of  $\sigma$ . The amplification (as compared to the SNR in the absence of added noise) can be quite large, particularly for small values of  $r(0)/f_c$  (ref. 6). Here we address a related but separate issue. With realistic biological parameters, can the addition of noise render detectable ( $\text{SNR} \geq 1$ ) a signal that without added noise has a SNR much less than unity?

For a biological cell with radius  $r_{\text{cell}}$  in an external electric field  $E_{\text{rms}}$  the dimensionless signal amplitude<sup>8</sup> is  $V_s = (zeE_{\text{rms}}r_{\text{cell}}/k_B T)$ , where  $z$  (a gating charge for an electrically sensitive protein) parametrizes the sensitivity of the biophysical detection mechanism to the applied field ( $k_B$  is the Boltzmann constant and  $T$  the temperature). The cut-off  $f_c$  is approximately the inverse of the charging time for the membrane (less than 10 MHz) so we use  $f_c = 10^7 \text{ s}^{-1}$ . In Fig. 1, a



**Figure 1** SNR against the  $\sigma$ , calculated using equation (1), for  $E_{\text{rms}} = 1 \text{ mV cm}^{-1}$  at room temperature ( $k_B T = 25 \text{ meV}$ ), and the parameters  $z = 10$  electronic charges,  $r_{\text{cell}} = 100 \text{ } \mu\text{m}$ ,  $f_c = 10^7 \text{ s}^{-1}$ ,  $r(0) = 10^6 \text{ s}^{-1}$ , and with  $\Delta f_a = 100 \text{ s}^{-1}$ , appropriate for a detector around 50–60 Hz (ref. 9).

plot of the actual numerical value of the SNR against noise amplitude for a specific set of parameters is shown for  $E_{\text{rms}} = 1 \text{ mV cm}^{-1}$ . Although we selected values to maximize the SNR within experimental constraints, the SNR does not approach unity.

To achieve a  $\text{SNR} > 1$  with  $r(0) < f_c$  (which is necessary to derive any benefit from added noise), the factor  $z^2/\Delta f_a$  would have to be much larger than 1. In principle, this could be achieved by a system with a tremendous sensitivity to an applied field ( $z > 10$ ), or by a detection mechanism that is sensitive only in a very narrow band of frequencies ( $\Delta f_a \ll 100 \text{ s}^{-1}$ ) centred around the signal frequency. However, in the absence of experimental evidence for such properties of cells, we conclude that, as for other mechanisms of single-cell stochastic resonance<sup>8</sup>, the mechanism of Bezrukov and Vodyanoy cannot explain experiments in which fields of less than  $1 \text{ mV cm}^{-1}$  are claimed to cause effects on single cells<sup>2</sup>.

#### R. Dean Astumian

Departments of Surgery and of Biochemistry and Molecular Biology,  
The University of Chicago,  
Chicago, Illinois 60637, USA  
e-mail: dastumia@surgey.bsd.uchicago.edu

#### R. K. Adair

Department of Physics,  
Yale University, New Haven,  
Connecticut 06520, USA

#### James C. Weaver

Harvard-MIT Division of Health Sciences and Technology,  
Massachusetts Institute of Technology,  
Cambridge, Massachusetts 02139, USA

**Bezrukov and Vodyanoy reply** — The possibility that the mechanism of stochastic resonance might account for the sensitivity of organisms to industrial low-frequency magnetic and/or electric fields has inspired much discussion<sup>1,2,10</sup>. In a recent Letter<sup>6</sup> we formulated a physical model of stochastic resonance that permits quantitative analysis of the phenomenon in a variety of sys-

tems. Astumian and co-workers apply our results to the intriguing but difficult question of the role of stochastic resonance in signal detection at the level of a single cell. For a 'typical cell' of  $100 \text{ } \mu\text{m}$  radius, they come to a conclusion that stochastic resonance increases cell sensitivity to small signals, but that this increase is not significant enough to account for the effects reported by others<sup>2</sup>. Although we agree with their treatment of our theory, we show that for smaller cells stochastic resonance can improve signal detection considerably.

Astumian and co-workers address two possibly related but altogether different questions: whether electrical power lines pose a significant health hazard and whether non-equilibrium noise can play a role in small-signal detection by single cells. Although the health-hazard question unfortunately cannot be answered by our physical (not physiological) theory, the question about a role of stochastic resonance in signal detection is in the domain of our model.

Here we would like to demonstrate that the choice of the system parameters is crucial. The vertical arrow in Fig. 2 shows the  $f_c/r(0)$  value used by Astumian and co-workers, which gives an increase in the cell sensitivity of about 1.8 times. However, consider a cell of 10 to  $30 \text{ } \mu\text{m}$  diameter. Using their reasoning that the corner frequency,  $f_c$ , is limited by cell membrane capacitance (that is, ignoring a possible frequency dependence of the channel response<sup>11</sup>) and taking into account that the number of channels (and thus  $r(0)$ ) is proportional to the membrane area, we obtain a increase in  $f_c/r(0)$  of two to four orders of magnitude. This increase in  $f_c/r(0)$  will correspond to a 10- to 100-fold improvement in signal-to-noise ratio, suggesting a significant role for stochastic resonance in signal detection. The r.m.s. value of optimal noise only doubles (Fig. 2). It can be shown that for  $f_c/r(0) < 10^5$ , the depen-

dence of optimal noise on this parameter is slower than logarithmic:

$$\sigma_{\text{opt}} \approx \sqrt{\ln \frac{f_c}{r(0)}} \quad (2)$$

Indeed, the decrease in the equilibrium rate  $r(0)$  will decrease the initial signal-to-noise ratio, but there is another parameter,  $\Delta f_a$ , that is as significant as  $r(0)$ , and that was arbitrarily chosen by Astumian *et al.* as  $10^2 \text{ s}^{-1}$ . In principle, the theoretical limit on  $\Delta f_a$  is defined by the device lifetime  $\tau$ , and in the case of a human being is  $\Delta f_{a \text{ lim}} = 1/(2\pi\tau) \approx 1/(2\pi \times 75 \text{ years}) \approx 4.2 \times 10^{-10} \text{ s}^{-1}$ . Individual cells on average do not live that long, so reducing the limiting frequency resolution to about  $10^{-8} \text{ s}^{-1}$ , but this is still quite different from  $10^2 \text{ s}^{-1}$ .

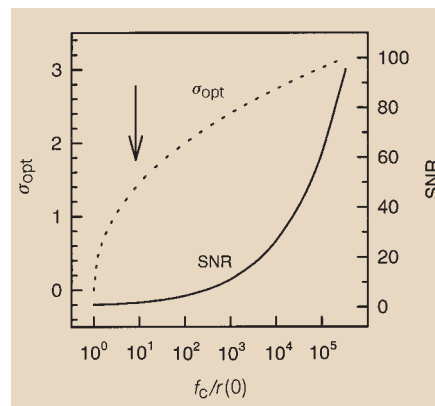
One of the main conclusions of our Letter<sup>6</sup> is that stochastic resonance is an inherent feature of a variety of systems with threshold-free behaviour. Indeed, a News and Views article in the same issue<sup>12</sup> mentioned a particular electronic device that can be described by our model, and predicted some other applications including biological ones. We were delighted to see yet another application made by Astumian *et al.* to describe stochastic resonance at the single-cell level. As for the health-hazard question mentioned above, resolving it will require joint efforts by physicists, physiologists and epidemiologists. What is clear at the moment is that specialized cells and organisms do have extraordinary electrical sensitivity<sup>13,14</sup>, and that some sensory biological systems and neuron circuits do show stochastic resonance<sup>15,16</sup>.

#### Sergey M. Bezrukov

Division of Computer Research and Technology,  
National Institutes of Health,  
Bethesda,  
Maryland 20892-0580, USA  
and St Petersburg Nuclear Physics Institute,  
Gatchina, 188350 Russia  
e-mail: bezrukov@helix.nih.gov

#### Igor Vodyanoy

Office of Naval Research, Europe,  
223 Old Marylebone Road,  
London NW1 5TH, UK



**Figure 2** Signal-to-noise ratio (SNR) and optimal noise r.m.s. amplitude  $\sigma_{\text{opt}}$  against the ratio of the noise bandwidth  $f_c$  to the equilibrium rate  $r(0)$ , calculated from the stochastic resonance model<sup>6</sup>.

- Krugilov, I. L. & Dertinger, H. *Bioelectromagnetics* **15**, 539–548 (1994).
- Kaiser, F. *Bioelectrochem. Bioenerget.* **41**, 3–18 (1996).
- Moss, F. & Wiesenfeld, K. *Nature* **373**, 33–36 (1995).
- Weaver, J. C. & Astumian, R. D. *Science* **247**, 459–462 (1990).
- Adair, R. K. *Phys. Rev. A* **43**, 1039–1048 (1991).
- Bezrukov, S. M. & Vodyanoy, I. *Nature* **385**, 319–321 (1997).
- Bezrukov, S. M. & Vodyanoy, I. *Nature* **378**, 362–364 (1995).
- Astumian, R. D., Adair, R. K. & Weaver, J. C. *Proc. Natl. Acad. Sci. USA* **92**, 3740–3743 (1995).
- Robertson, B. & Astumian, R. D. *J. Chem. Phys.* **99**, 7414–7419 (1991).
- Adair, R. K. *Bioelectromagnetics* **17**, 242–245 (1996).
- Bezrukov, S. M. & Vodyanoy, I. *Biophys. J.* (in the press).
- Jung, P. & Wiesenfeld, K. *Nature* **385**, 291 (1997).
- Kalmijn, A. D. *Science* **218**, 916–918 (1982).
- Lu, J. & Fishman, H. M. *Biophys. J.* **67**, 1525–1533 (1994).
- Bulsara, A. R. & Gammaitoni, L. *Phys. Today* **49**, 39–45 (1996).
- Douglas, J. L., Wilkens, L., Pantazelou, E. & Moss, F. *Nature* **365**, 337–340 (1993).



# Sequence-specific RNA binding by Bicoid

It has recently been reported that the *bicoid* protein (Bcd), a homeodomain-containing protein which forms an anterior-to-posterior gradient during early *Drosophila* embryogenesis, can bind messenger RNA encoding the *caudal* protein (Cad) and repress its translation<sup>1,2</sup>. This may be mediated by direct interactions between the homeodomain and a discrete RNA sequence in the untranslated 3' region of the *cad* transcript, the Bcd response element (BRE)<sup>1,2</sup>. The evidence for direct and specific homeodomain-RNA interac-

tions depends principally on the results of competition experiments<sup>1</sup> but these could not be reproduced under the conditions originally reported<sup>3</sup>. Here we present competition experiments performed under different conditions confirming that direct, sequence-specific interactions can occur between the Bcd homeodomain and the *cad* BRE. We also extend previous evidence<sup>1</sup> that site-specific RNA binding by the Bcd homeodomain depends on the lysine at position 9 of helix III, an amino-acid that mediates site-specific DNA binding<sup>4-6</sup>.

We expressed homeodomain peptides from *bcd* and *Ultrabithorax* (*Ubx*) as glutathione-S-transferase conjugates in *Escherichia coli*. After purification, we assayed their ability to bind to radio-

labelled RNAs comprising either the *cad* BRE or the 3' untranslated region of the *Tubulin1* (*Tuba1*) gene using band-shift experiments.

The BRE RNA, but not the *Tuba1* RNA, seems to be bound by the Bcd homeodomain as indicated by the shift in mobility of the RNA. This suggests that an RNA-protein complex is formed (Fig. 1a). Formation of this complex is severely inhibited by a 10-fold excess of cold competitor BRE and we could no longer detect it in the presence of a 100-fold excess (Fig. 1a). In contrast, formation of the complex is not severely inhibited by a 100-fold excess of cold *Tuba1* RNA or yeast transfer RNA (Fig. 1a).

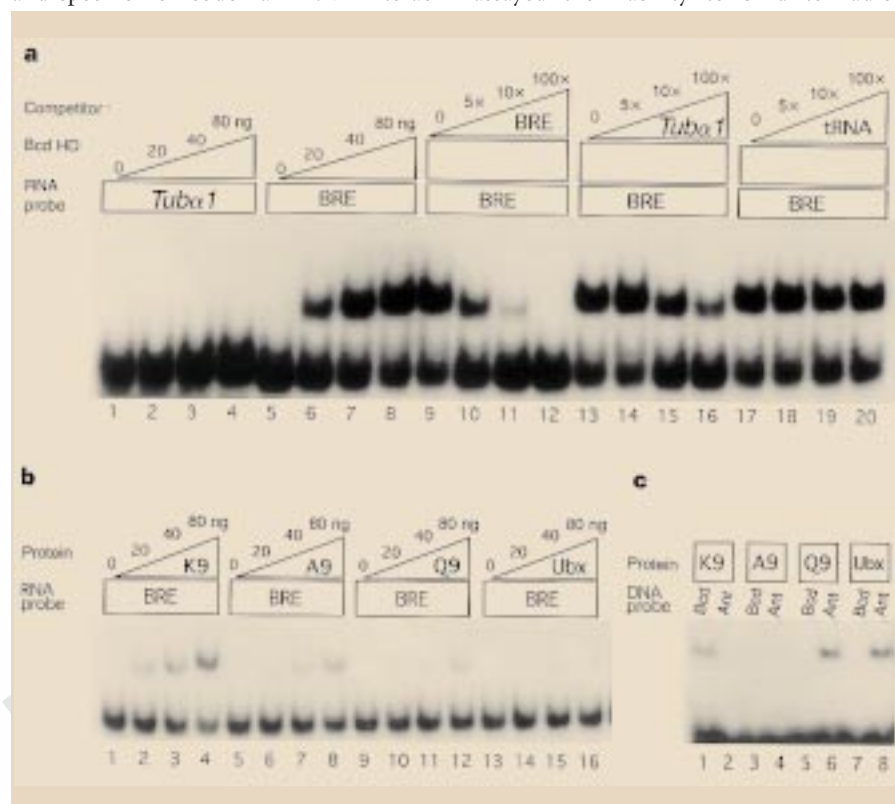
Substitution of the lysine at position 9 of the Bcd homeodomain (K9) by alanine (A9) or glutamine (Q9) causes a fivefold or 10-fold reduction, respectively, in the amount of the RNA-protein complex formed (Fig. 1b), correlating with the failure of full-length Bcd protein carrying these point mutations to repress Cad expression *in vivo*<sup>1</sup>. Virtually no such complex forms after incubation with a *Ubx* homeodomain peptide, which contains glutamine at position 9 but differs from the Bcd<sup>Q9</sup> peptide at 36 of the remaining 61 residues. As expected from studies<sup>4-6</sup>, the Bcd<sup>Q9</sup> and Bcd<sup>A9</sup> homeodomains have altered DNA-binding specificity (Fig. 1c): the Bcd<sup>Q9</sup> peptide, like the corresponding *Ubx* peptide, binds Antp rather than Bcd target sequences, whereas the Bcd<sup>A9</sup> peptide binds neither.

We conclude that the lysine at position 9 in the Bcd homeodomain is important in mediating site-specific RNA and DNA binding by the intact protein.

Siu-Kwong Chan

Gary Struhl

Howard Hughes Medical Institute,  
Columbia University College of Physicians  
and Surgeons,  
701 West 168th Street,  
New York, New York 10032, USA  
e-mail: struhl@cuccf.ccc.columbia.edu



**Figure 1a**, Radiolabelled BRE or *Tuba1* 3' UTR RNAs were incubated with different amounts of purified glutathione-S-transferase Bcd homeodomain (HD), with or without cold competitor RNAs as indicated. Radiolabelled BRE RNA, but not *Tuba1* RNA, is bound by the Bcd homeodomain peptide. Binding of the BRE RNA is competed roughly 10 times more efficiently by cold BRE than cold *Tuba1* competitor RNA (average of three independent experiments) indicating that it is sequence-specific. **b**, Radiolabelled BRE RNA was incubated with purified glutathione-S-transferase conjugates of the homeodomains of wild-type Bcd (K9), either of two Bcd mutants (A9, Q9), or *Ubx*. Formation of RNA-protein complexes was analysed as before. The A9, Q9 and *Ubx* peptides form roughly fivefold, 10-fold and 100-fold less BRE-protein complex, respectively, relative to the wild-type (K9) peptide (average of three experiments). **c**, Purified glutathione-S-transferase conjugates (5 ng) of the K9, A9 and Q9 derivatives of the Bcd or *Ubx* homeodomains were incubated with radiolabelled DNA duplexes containing either the Bcd-class or Antp-class (Ant) binding sites (cgtagctc-TAATCCGggcgcg and cgtagctc-TAATTAGggcgcg, respectively)<sup>4-6</sup> and binding was analysed using a band-shift assay. K9, Q9, and *Ubx* peptides bind selectively to their appropriate targets (Bcd-class for K9; Antp-class for Q9 and *Ubx*). A9 peptide binds to neither target. RNA-binding conditions: 10 mM HEPES buffer at pH 7.5, 100 mM KCl, 1 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 1 mM DTT, 2.5 µg µl<sup>-1</sup> heparin, 0.1 µg µl<sup>-1</sup> yeast tRNA, 20 units RNase inhibitor, 10% glycerol, incubated for 20 min at 20 °C. Radiolabelled BRE and *Tuba1* RNAs contain nucleotides 210 to 253 of the *cad* 3' UTR and -10 to 307 of *Tuba1* 3' UTR, respectively<sup>1</sup>. Gel-purified RNA (1 ng) with specific activity of ~10,000 c.p.m. per ng was used in each binding reaction. Further details are available on request from the authors.

1. Dubnau, J. & Struhl, G. *Nature* **379**, 694-699 (1996).
2. Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J. & Jäckle, H. *Nature* **379**, 746-749 (1996).
3. Dubnau, J. & Struhl, G. *Nature* **388**, 697 (1997).
4. Hanes, S. D. & Brent, R. *Cell* **57**, 1275-1283 (1989).
5. Treisman, J., Gönczy, P., Vashishtha, M., Harris, E. & Desplan, C. *Cell* **59**, 553-562 (1989).
6. Hanes, S. D., Riddihough, G., Ish-Horowitz, D. & Brent, R. *Mol. Cell Biol.* **14**, 3364-3375 (1994).

## correction

In 'Metallothionein in snail Cd and Cu metabolism' by R. Dallinger *et al.* (*Nature* **388**, 237-238; 1997) the relative molecular masses of the metallothioneins were given incorrectly. The correct relative molecular mass of the metallothionein isoform found in the midgut gland is 6,620 (6.62K) and that of the isoform found in the mantle is 6,247 (6.247K).

# I'll be judge, I'll be jury

## Judging Science: Scientific Knowledge and the Federal Courts

by Kenneth R. Foster and Peter W. Huber  
MIT Press: 1997. Pp. 333. \$40, £33.95

Sheila Jasanoff

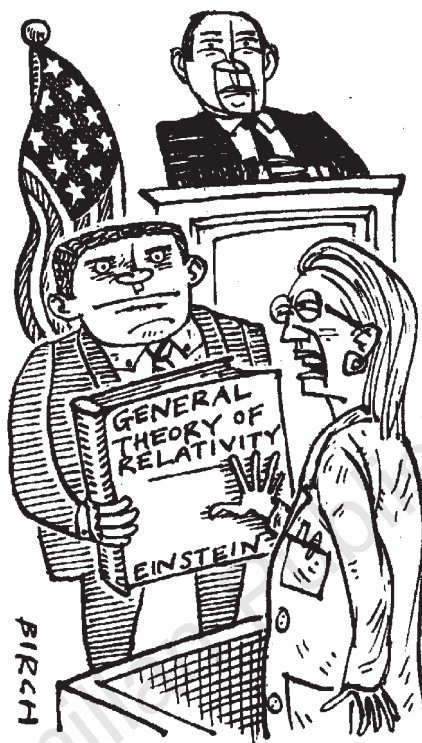
In an earlier book on science and litigation (*Galileo's Revenge*, BasicBooks, 1991), Peter W. Huber, a lawyer and fellow of the conservative Manhattan Institute for Policy Research in the United States, made sound-bite history with the phrase 'junk science'. An instant hit with journalists, politicians, corporate lawyers and many scientists, it resonated well with the across-the-board retreat in the United States from populism, judicial activism and liberal permissiveness.

Huber used the term to lampoon the legal system's supposed inability to distinguish scientific sense from nonsense. Suddenly, there was a simple explanation for everything the matter with courts. Technically illiterate judges and juries could be blamed for a host of ills, from excessive liability awards to perceived economic and technological stagnation. In 1994, the Republican party platform officially embraced the mission to make the law more accountable to mainstream science.

Those looking to this book for an encore may be disappointed. Gone for the most part are Huber's amusing and politically effective verbal pyrotechnics. A more measured tone prevails — a result, perhaps, of the collaboration with Kenneth R. Foster, an engineer who has for several years partnered Huber in the project to improve the quality of the science relied on by US courts.

Times have changed in other ways, too. In 1993, the US Supreme Court decided a case, *Daubert v. Merrell Dow Pharmaceuticals Inc.*, that encouraged federal judges to act more aggressively as gatekeepers of science. Foster and Huber want to ensure that *Daubert* will be interpreted so as to raise to new heights the threshold for admitting scientific evidence. Their aim is superficially narrow: to explicate in detail the key terms used in the *Daubert* opinion, such as scientific validity, reliability and 'fit'. On the face of it, the book looks like a specialist essay on a technical topic addressed to a professional legal audience.

Appearances, however, can be profoundly deceptive. If the authors had their way, there would be a radical restructuring of power and authority in US courtrooms, possibly undermining the constitutional right to a jury trial. In Foster and Huber's scheme, nothing short of peer-reviewed, published scientific findings would be admitted into evidence, and preference would be given to consensus statements by distinguished expert bodies such as



the National Research Council.

The exclusion of less certain evidence would eliminate much of what today passes as fact-finding in legal proceedings. Juries would become an endangered species, at least in cases involving complex arguments about disease causation. Ironically, the very judges whose scientific competence Huber once mercilessly impugned would emerge, in the post-*Daubert* era, as capable of the greatest technical discernment.

There is a misleading seductiveness to the argument. Throughout the book, the authors, quite sensibly, call attention to the provisional quality of science; the absence of any fixed indicators of scientific validity; and the similarities between scientific and other forms of reasoning. Their expectations of the legal system at times sound similarly reasonable.

For example, Foster and Huber agree with the *Daubert* majority that the goals of the law are not to achieve "cosmic understanding" but "to arrive at a reasonably reliable answer in a reasonably short period of time". The implication is that courts should settle for less-than-perfect scientific knowledge. It is only in piecing together what Foster and Huber mean by "reasonable" and "reliable" that the far-more-controversial character of their message becomes clear.

Using Bayesian statistics as a rationale, the authors imply that courts should not have compensated asbestos workers for mesothelioma in the 1950s; not until the 1970s did the data converge on a reliable showing of causation, they feel. They dismiss

the findings of cancer risk from passive smoking as being "at best — at the very edge of detectability". Reports of disease clusters are seen as worse than useless, leading more often to unjustified panic than to valid claims for compensation.

Replication subtly becomes the touchstone that courts should look for in evaluating science. Reliability, the authors say, develops only when "other scientists pick up the trail and independently investigate phenomena and test theories". Following this reasoning, the standard of admissibility for scientific evidence creeps perilously close to complete certainty, a conclusion patently inconsistent with the *Daubert* majority opinion.

Slippery formulations abound in the book, not only in statements of the authors' own positions but also in their representation of the scholarship of others. Research in science and technology studies, for example, is erroneously characterized as dealing only with the "external" aspects of science: that is, with people and not with the content of their ideas.

More problematically, Foster and Huber attribute to the sociology of science a mindless relativism that betrays a serious misreading of the few works they cite as well as of the field's contributions to elucidating the construction of scientific knowledge. Political motives are casually imputed to others in a work that seems thoroughly unselfconscious about its own political affiliations.

A surprising amount of space is devoted to excerpts from other authors who have written on science, law and evidence. These boxed and edited passages, often more than a page in length, detract irritatingly from the coherence of the authors' own presentation. Offered with little interpretive guidance, the extracts create a scrapbook of oddly unassembled voices that is more appropriate to a postmodern experiment with novel textual forms than to sober analytical reasoning.

In its single-minded emphasis on making US courtrooms safe for science, the book overlooks crucial issues of law and public policy. It lacks any explicit recognition of the jury's sovereign right to 'find facts' or of the dilemmas this right entails for judges, who daily grapple with the fuzzy boundary between fact and law. The policy implications of allowing unelected judges to exclude evidence, often on the basis of an incomplete record, remain unexplored. And the tragic question of how justice can be rendered in a world of irreducibly uncertain knowledge is bypassed in silence. □

Sheila Jasanoff is in the Department of Science and Technology Studies, Cornell University, 632 Clark Hall, Ithaca, New York 14853, USA.

# On the horns of a dilemma

## Horn of Darkness: Rhinos on the Edge

by Carol Cunningham and Joel Berger  
Oxford University Press: 1997. Pp. 246.  
£18, \$25

Brian Bertram

Wild rhinoceros populations are plummeting — precipitously or just fast, depending on the timescale one chooses. Rhinos were once widespread across Africa and Eurasia, with their remains abundant as fossils and their likenesses represented in cave art from many early human societies.

Half a century ago there may have been 100,000 black rhinos in Africa, and a mere 20 years ago the population was at least 10 times its present estimated figure of 2,000. Over the past couple of years, the actual drop in numbers may not have been very large, but the outlook for the rhino is grim.

A small part of the crash has, as usual, been due to the appropriation of rhino habitat by soaring human populations. But the main problem has been the widespread deliberate illegal slaughter of rhinos for their horns, traded internationally. There have been surging and fluctuating demands for rhino horn for various medicinal purposes in the Far East, and as dagger handles in North Yemen; increasing wealth in these territories has resulted in end market prices of US\$3,700 to US\$16,000 per kilogram for rhino horn. There is obviously a huge financial incentive to poach rhinos.

Efforts to combat rhino poaching in Africa, with paramilitary operations, were largely unsuccessful. This is hardly surprising as the fight was against well-equipped poaching gangs over huge areas. Also, even if it is practicable, many people are uncomfortable with the idea of guards shooting people who are shooting rhinos.

One solution tried in Namibia in the early 1990s was to de-horn rhinos. In principle, without its valuable horn, a rhino would no longer be a potential target for poachers. There have been hotly debated questions surrounding the issue. The biological

questions are: do wild rhinos need their horns, for instance to defend themselves or their young against predators? And how fast do the horns re-grow in the wild? The cultural questions are: do poachers actually discriminate, or would they shoot any rhino regardless? The economic questions are: might cut-off horns bring in good revenue for government when or if legal international trade in rhino horn is resumed? Are overseas funds more readily available for rhino conservation measures that involve the practical approach of de-horning? And might future rhino-hunting safaris to de-horn rather than kill the quarry provide good foreign exchange?

Clearly the biological questions need to be answered by research, hence Carol Cunningham and Joel Berger's study: to look at the effects of de-horning by comparing horned and de-horned rhinos in terms of reproduction, survival and behaviour. Additional information was to come from studying the natural variation in horn size and the growth rates of horns.

It sounds straightforward. In practice it is incredibly difficult, and to have obtained any information at all is a triumph.

In the Namib Desert and in Etosha, rhinos are sparse, elusive and inaccessible (or they would have been poached out before now). Studying them consisted largely of waiting in a Land Rover near waterholes until after dark, when rhinos might appear, then tiptoeing on foot until within camera range and trying to get flash photographs good enough for measurement of the horns and identification of the individual.

The determined, gutsy couple spent 197 nights with rhinos, with 1,030 hours of observation of about 100 known individuals — a truly tremendous achievement, for which they deserve huge credit.

The variety of the task, spread over three field seasons, was added to by the company of Cunningham and Berger's baby daughter and of Huie (alias Archie) Gawuseb, their Damara tracker; by encounters with lions and elephants; by the trials and tribulations of keeping a Land Rover going in the desert; and by the apparent resentment of some Namibian government conservation staff towards them as foreign researchers. This resulted in their fourth visit being cut off by revocation of their research permit.

Their book is full of variety, if not always clarity. It is written in a bitty, curiously innuendo-rich manner. Often it seems there is something to read between the lines, but usually it is not clear what. The main conclusions from the study, as reported in the book, would appear to be that black rhino horns re-grow at a rate of about half a kilogram per year, that poachers probably do not discriminate in what rhinos they kill, and that hornless rhinos may possibly suffer higher calf mortality, perhaps because of hyena predation.

The answer to the question "Does rhino de-horning work?" is a gloomy one. So far, almost nothing has been shown to work that will prevent these magnificent animals being needlessly slaughtered. What will our grandchildren think of us all? Cunningham and Berger can at least show that they tried, and did so valiantly. □

Brian Bertram is at Amberley, Stroud, Gloucestershire GL5 5AG, UK.

# Motion of the heavens

## Celestial Encounters: The Origins of Chaos and Stability

by Florin Diacu and Philip Holmes  
Princeton University Press: 1997. Pp. 233.  
\$24.95, £19.95

Jacques Laskar

Chaos has recently become fashionable, with the appearance of several popular books on the subject, including bestsellers such as James Gleick's *Chaos* (1989). The theory of chaos is often presented as a new science that began with the numerical simulation of dynamical systems by computers. Florin Diacu and Philip Holmes have written their book partly in reaction to this trend. They deliberately omit findings obtained with the help of computers, and concentrate instead on mathematical results.

Written for a general audience, the book begins with the exciting story of Henri Poincaré's discovery a century ago of the 'homo-





clinic tangle', which lies at the root of the chaotic behaviour of orbits. In his original memoir on the three-body problem, which in 1889 won a prize sponsored by King Oscar II of Sweden and Norway, Poincaré failed to understand properly the complexity of the trajectories. It was only after publication of his memoir in *Acta Mathematica* that he realized his mistake. He had to have printed copies of the journal withdrawn to correct his description of the homoclinic tangle.

The authors next consider the question raised by Paul Painlevé in 1897 about the possible existence of singularities in the  $n$ -body problem that are not due to collision. This well-formulated problem had to wait 100 years for its resolution through the recent solutions by Gerver and Xia in which bodies travel an infinite distance in finite time. These solutions are of a purely mathematical nature and cannot exist physically, as they require dimensionless bodies and arbitrarily close encounters.

Whereas Painlevé's conjecture is purely mathematical, the question of the stability of the Solar System is of a different nature. Although the problem can be formulated mathematically — as Poincaré did in asking for the infinite time stability of a point-mass planetary system — the real question of interest is more the practical stability of the Solar System over periods of time comparable to its age. Diacu and Holmes treat this question less exhaustively than they do that of collisionless singularities.

Laplace and Lagrange showed between 1799 and 1825 that, as a first-order approximation with respect to the planetary masses, the semi-major axes of the planets do not undergo long-term — or secular — variations. However, this fact is not sufficient to prevent planetary collisions, because even if the semi-major axes of the planets are constant, changes in their eccentricities can induce their orbits to cross.

To ensure the stability of the Solar System,

it is also necessary to provide bounds on the variations of the planetary eccentricities. This task was again carried out by Laplace and Lagrange, who demonstrated the linear stability of the first-order secular equations of the eccentricities and inclinations.

It is surprising that this part of the story is omitted by the authors, for on this topic Laplace gave a brilliant mathematical proof. The authors, to my mind, underestimate the fundamental value of these results, saying they would "refer to the  $n$ -body problem and not directly to the Solar System itself". In fact, Laplace's solution — which was fully computed for the whole Solar System by Le Verrier in 1856 — provides a good approximation to the behaviour of the Solar System over a few million years.

The next natural question concerns the modifications that might arise out of the higher-order terms not taken into account in Laplace's computation. Attempts to answer this question formed a quest lasting nearly

## Film review Aliens, lies and videotape

### Contact

A film by directed by Robert Zemeckis and co-produced by Carl Sagan and Ann Druyan  
Warner: 1997

*Contact* is based on a 1995 novel by Carl Sagan, the astronomer and popularizer of science who died last year. The story's central character, Ellie Arroway (Jodie Foster) is also an astronomer, and devotes her career to searching for radio signals from civilizations on planets orbiting other stars. When she succeeds in detecting a signal, she is propelled into a political controversy for which she is ill-prepared.

The decoded message turns out to be instructions for building a large device, the ultimate purpose of which is unclear. In the last half hour of the film, Ellie tests the device, makes contact with the aliens, but carries back to Earth no objective proof (namely on video) of her experience.

Up until the detection of the signal, this entertaining and thought-provoking film is an accurate portrayal of a radioastronomer's life, showing that scientists, just like other people, are driven by complex emotions. And the main subplot of a senior scientist taking over — and taking all the credit for — a project that has suddenly become successful is sadly not unknown, especially in astronomy.

One early theme is that Arroway is committing professional suicide by her involvement in the search for extraterrestrial intelligence (SETI). Young astronomers who have not yet obtained tenure will recognize the truth in this portrayal. In essence, SETI is a scientific lottery: the winner will certainly achieve scientific immortality, but will anyone win?

I suspect few astronomers would say on the record that life elsewhere in the Universe does not exist. But communicating with another



civilization is a different matter. One of the basic problems is the length of time a technological civilization such as our own would survive; another is the duration of the period in which an advanced civilization would be interested in trying to communicate with other worlds. Where should we look? Which stars should we examine? At what what frequencies should we look for a signal? Finally, there is the question of language.

In *Contact*, the issues of frequency and language are handled deftly, although perhaps cryptically for many lay people. Arroway exclaims "pi times H" upon finding the signal, which all astronomers will immediately interpret as  $\pi \times 1,420.406 \text{ MHz}$  — 1,420.406 MHz, being the frequency of a photon emitted as a hydrogen atom, makes a transition between two hyperfine levels in the ground state. The implicit idea is that any technological civilization would know both numbers; this sets the stage for mathematics and science being the universal language.

Refreshingly, the science and terminology adopted in the film are accurate, even though some artistic liberties are taken. (For example, the Very Large Array (VLA) is not located within walking distance of Canyon de Chelly, nor can

you plug headphones into the base of a VLA antenna.) Great care has been taken to get the details right. Although some lay people might find the jargon intimidating, it is enjoyable to hear it used properly, and those unable to follow the details will not miss anything essential to the story. All too often, the level of science in films tends to be abysmal, with simple words or concepts mangled beyond recognition.

The wider issue of whether we should be engaged in SETI, and how much effort should be devoted to the search, is handled in an impassioned speech from Arroway as she seeks private funding after the US National Science Foundation terminates her project. This is pure Sagan, right down to the turtleneck sweater and tweed jacket Arroway wears. One wonders whether he regretted not having shown as much passion as she does here when all NASA funding for SETI was eliminated by Congress in 1993.

The question "Are we alone?" has long intrigued humanity. "What a waste of space if we are alone," say several of the film's characters.

*Contact* conveys, eloquently and vibrantly, Sagan's belief that we are not.

Leslie Sage is an assistant editor of *Nature*.

100 years, until Poincaré demonstrated the divergence of the series used by astronomers.

The work of Poisson and Haretu belongs to this quest. Their goal was to compute the higher-order terms in the semi-major axis variations. Poisson demonstrated that, as a second-order approximation with respect to the masses, the semi-major axes do not present purely secular terms but instead have mixed terms. Then Haretu showed in his thesis the possibility of secular terms to third order with respect to the masses.

I find it hard to support the authors' enthusiastic commentary on Haretu's work: "Haretu's theorem can be considered as a foundation stone for the great edifice of KAM theory". Unlike Poisson's results, which inspired Poincaré's famous recurrence theorem, Haretu's results do not appear to have had any profound repercussions. And as a response to a decades-old question, the implications of his admittedly important findings are limited: the secular terms found by Haretu are simply the Taylor expansion of some small periodic terms due to the quasiperiodic variations of the eccentricities as they appear in Laplace's solutions. The possibility of instabilities in the Solar System does not result from Haretu's work but from Poincaré's demonstration of the divergence of the series of celestial mechanics.

Towards the end of the book, the authors suggest that Xia's recent results on Arnold diffusion in the three-body problem could help to explain the Kirkwood gaps in the asteroid belt. I cannot agree.

First of all, the authors seem to forget that recent extensive work using numerical techniques and perturbation theory now provides a virtually complete view of the dynamics of the asteroid belt. It is true that these results are not "rigorous mathematical proofs", but then again the Solar System is not a mathematical object. Moreover, most of these results can be related locally to small models for which the dynamical behaviour is relatively well understood. It is therefore unfair to disregard these works.

On the other hand, Xia's results apply only to a caricature of the Solar System, require unrealistically small perturbations and result in extremely bad time-scales for the diffusion. Although these results (if vindicated mathematically — the proofs are disputed) may be a remarkable mathematical achievement, they seem to provide no better insight into the physical understanding of diffusion in the asteroid belt than does Arnold's original result of 1964.

Diacu and Holmes confront a difficult challenge in aiming to provide an accessible account of celestial mechanics as the origin of many of the ideas in chaos theory. In doing so, they seem driven to taking certain liberties with the strictly historical account. In particular, many of the story's characters are given thoughts they probably never had, and

are generally portrayed as behaving like modern US scientists.

Despite these shortcomings and some obvious subjectivity in the narration, this book is welcome for its detailed account of the resolution of Painlevé's conjecture. It will provide the reader with a lively introduction to the fascinating story of celestial mechanics. □

Jacques Laskar is at *Astronomie et Systèmes Dynamiques, Bureau des Longitudes, 3 rue Mazarine, F-75006 Paris, France.*

## Outdoor man

### **Eco Homo: How the Human Being Emerged from the Cataclysmic History of the Earth**

by Noel T. Boaz

*BasicBooks: 1997. Pp. 278. \$25*

### **The Aquatic Ape Hypothesis**

Elaine Morgan

*Souvenir Press: 1997. Pp. 205. £16.99*

Ian Tattersall

One of the trendiest recent exercises in palaeoanthropology has been the publication of books that in one way or another blame the emergence and subsequent evolution of our own dubious hominid lineage on the climatic oscillations of the Pleistocene epoch (the 'ice ages') and earlier.

So far, so good; for although some would argue long and hard about claimed correlations of climatic and ecological events with particular innovations in the hominid record, I think I'm in pretty good company in finding it inconceivable that the dramatic worldwide warmings and coolings of the Pleistocene, and the sea-level (hence geographic) changes that accompanied them, did not radically affect the course of hominid evolution. What I find truly remarkable, though, is how little these books have in common beyond the shared conviction that human evolution over the past two or three million years has been driven by the vicissitudes of climate.

Noel T. Boaz's latest offering is a case in point. Most authors on this subject find a congenial fit between a geological record of unstable climates and geographies and the notion of evolution by punctuated equilibria ('fits and starts'). Boaz, however, explicitly dismisses this evolutionary model on the rather odd and unexplained grounds of 'vitalism' and a 'lack of clear causative association in nature'. He prefers an older model of continuous evolution with varying rates, the background rate dictated by the 'Red Queen' model ('running' to stay in the same place owing to competition) and faster change coming at times of dramatic climatic shift. Unfortunately, this choice doesn't help his argument: the Pleistocene record is one of almost ceaseless environmental oscillation, so his strictly adaptationist perspective forces him to seek many more evolutionary

correlates of climatic change than other authors have been comfortable with. What is more, the choice may also account for his curious neglect of the major lineage-independent 'turnover pulses' that are a central feature of the faunal history of the Plio-Pleistocene.

Boaz covers a lot of ground, beginning his story in the Miocene, well before the origin of hominids, and continuing through the first australopiths and early *Homo* to the first Europeans and Asians and eventually to the rise and demise of the Neanderthals and their replacement by modern humans. Although selective, this is a competent enough survey. But it shows how hard it is for Boaz to reconcile his preferred evolutionary model with his desire to tie in virtually all human evolutionary developments with ecological cataclysms. In turn, his emphasis on the primacy of ecological change collides with his reluctance to recognize more than the smallest possible number of extinct hominid species at a time when conditions for speciation could not have been more propitious. These tensions are never resolved.

Elaine Morgan is another strict adaptationist who should, in theory, get along well with Boaz. At least until she reads in his book that humans are not well adapted for swimming. Morgan, as is well known, is the foremost if not the only active proponent of the 'aquatic ape theory', which posits human descent from at least a semi-aquatic ancestry. Over the decades since *The Descent of Woman* was published, Morgan has honed her writing and diplomatic skills, and this fourth book in her series is an engaging presentation of a scaled-down hypothesis that simply asks that we should seriously consider that a few million years of at least semi-water-adapted ancestry might be responsible for some of our anatomical features.

Only Morgan acolytes will warm to this appeal; but along the way she takes liberal swipes (some better aimed than others) at a whole range of more orthodox views, and clearly has a good time doing so. As she has noticed, palaeoanthropologists are fond of telling each other 'Just-So' stories; and once in a while a little needling of this kind does no harm at all. □

Ian Tattersall is in the Department of Anthropology, American Museum of Natural History, New York, New York 10024, USA.

The second edition of *Bones of Contention: Controversies in the Search for Human Origins* by Roger Lewin is now available in paperback. In a new afterword, the author looks at the way in which palaeoanthropology, while becoming more scientific, remains contentious. "Lewin may have contributed more to the evolution of a science than he had dared hope," wrote C. K. Brain in a review in *Nature* (330, 277; 1987). University of Chicago Press, \$18.95, £15.25.

# Fossils, genes and the evolution of animal limbs

Neil Shubin, Cliff Tabin & Sean Carroll

**The morphological and functional evolution of appendages has played a crucial role in the adaptive radiation of tetrapods, arthropods and winged insects. The origin and diversification of fins, wings and other structures, long a focus of palaeontology, can now be approached through developmental genetics. Modifications of appendage number and architecture in each phylum are correlated with regulatory changes in specific patterning genes. Although their respective evolutionary histories are unique, vertebrate, insect and other animal appendages are organized by a similar genetic regulatory system that may have been established in a common ancestor.**

The origin of evolutionary novelties raises some of the most fundamental questions of biology. How do new structures arise? Can they evolve *de novo* or are they generally derived from pre-existing structures? And what is the developmental and genetic basis for their origin and modification?

The adaptive evolution of vertebrates and arthropods to aquatic, terrestrial and aerial environments was accomplished by the invention of many novel features, especially new types of appendages. Enormous progress has been made in the past few years in understanding appendage development in both phyla. These genetic discoveries can be integrated with palaeontological data to address some of the principal events in the history of animal designs.

We will first examine the origin and evolution of vertebrate limbs and digits and of arthropod legs and insect wings. In both phyla we are confronted with a similar issue, namely the origin and adaptive modification of serially homologous organs. We will integrate palaeontological and developmental evidence that suggests that major innovations are largely derived from pre-existing developmental systems and will illustrate the potential genetic regulatory changes that enabled appendage evolution. Then we will explore the significance of newly discovered genetic similarities between arthropod and vertebrate appendages—similarities that have been retained despite more than 500 million years (Myr) of independent evolution. We will develop the hypothesis that the evolution of successively derived limb types, from lobopods to insect wings, and from agnathan fins to tetrapod limbs appears to be due, in part, to the successive cooption and redeployment of signals established in primitive metazoans. These examples illustrate how comparative developmental genetics can provide a mechanistic explanation of the origin and evolution of structures when palaeontological data are robust and important new hypotheses about evolutionary history when the fossil record is silent.

## Origin and diversification of tetrapod limbs

Vertebrate limb diversity was produced by changes in the number, position and shape of structures that can be traced to Ordovician<sup>2,3</sup> (463–439 Myr) through Late Devonian<sup>2–4</sup> (409–362 Myr) fossils. The demands of feeding and locomotion in Ordovician and Silurian seas led to a surprising variability of the earliest known appendages: some forms possessed a continuous anterior fin that ran the length of the body, others had paired fins that projected immediately behind a head shield, and still other primitive vertebrates had no paired fins at all (Fig. 1). A body plan with two sets of paired appendages, pectoral and pelvic, is a derived feature that first appears in later jawed vertebrates (gnathostomes)<sup>2–4</sup>. The number of paired appendages has been highly conserved ever since their origin: the evolution of new gnathostome body plans primarily involved a modification of existing paired appendages rather than the invention of whole new sets (acanthodians are the only excep-

tion to this generalization). Therefore, the origin of more recent novelties, such as digits, involved the modification of genetic systems first established in more primitive vertebrates.

## Serial homology and adaptive diversification

Primitive genetic systems must have provided a framework for the evolutionary integration of pectoral and pelvic appendages. Digits, for example, arose at the same time in the hand and foot: there is no Devonian tetrapod that has fingers and no toes<sup>2–5</sup>. Even in the post-Devonian world many unique designs appeared simultaneously in forelimbs and hindlimbs, as witnessed by chameleons, ungulates and ichthyosaurs. Obviously, serially homologous appendages can also evolve independently, an extreme case being the modification of pectoral appendages into wings in bats, birds and pterosaurs. Even in these extremely modified forelimbs, however, numerous similarities are retained between wing and leg.

The linkage between forelimbs and hindlimbs appears to be an ancient feature that resulted from patterns of gene cooption during the evolution of Palaeozoic fish. Serially homologous paired appendages are seen in Palaeozoic placoderms, acanthodians, chondrichthyans and osteichthyans<sup>2–4</sup>. In addition, pectoral and pelvic fins have evolved in parallel in almost all major gnathostome clades. There are numerous genetic parallels in pectoral and pelvic development that could account for these patterns of concerted evolution<sup>6,7</sup>. *Hox* genes, in particular, are likely to have been involved in the evolution of serial homology<sup>8</sup>. The earliest vertebrate appendages (unpaired fins) presumably did not utilize *Hox* genes; *HoxA* and *HoxD* genes are not expressed during the outgrowth of zebrafish unpaired fins<sup>9</sup>. Although these *Hox* genes were probably not involved in the origin of outgrowths in basal vertebrates, their later recruitment in the development of paired appendages was a key step in establishing serially homologous designs. The *HoxD* genes that came to play a role in appendage development are a subset of those involved in specifying regional identities along the caudal body axis (caudal neural tube, gut, somitic and lateral plate mesoderm)<sup>6–10</sup>. This suggests one of two situations: either nested patterns of *Hox* expression were originally present in a caudal set of paired outgrowths, and were later recruited in the development of a cranial set of outgrowths<sup>8</sup>, or similar *Hox* genes were recruited in pectoral and pelvic outgrowths at the same time in the evolutionary history of vertebrates (Fig. 1). In either case, pectoral and pelvic appendages were genetically linked early in their history and could have evolved together, presumably because the development of these appendages had already been brought under similar regulatory controls.

Superimposed on these ancient genetic parallels are secondary differences in gene expression and interaction that may have served as the basis for the independent evolution of pectoral and pelvic appendages. *Hox* gene expression in extant tetrapod limbs is dynamic and encompasses at least three distinct phases initiated

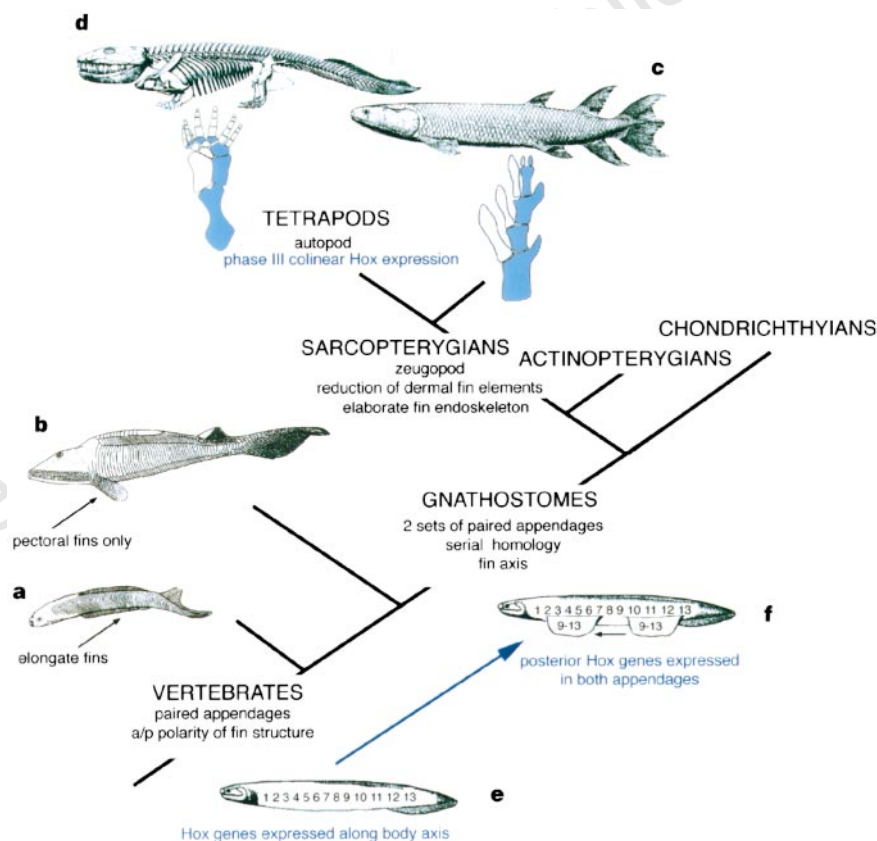


successively in the primordia of the stylopod, zeugopod and autopod<sup>7</sup> (Fig. 2). The presence of three distinct phases of *Hox* expression in limbs may reflect the observation that all tetrapods maintain a standard pattern of organization (Fig. 2), whereas specific differences in expression (or gene interaction) in each phase could result in the independent modification of pectoral and pelvic appendages<sup>7</sup>. Phase II *Hox* expression is practically identical in the fore- and hindlimb buds of mice, which possess generally similar skeletal patterns. In contrast, the wings and legs of chicks have very different skeletal patterns and different patterns of phase II *Hox* expression as well<sup>7,11</sup>. Surprisingly, phase III expression is very similar in chick wing and leg buds, indicating that aspects of the derived structure of chick wings are established by some other genetic means. Candidates include *Hox* genes of other clusters that are differentially expressed in the wing and leg buds<sup>12</sup> and the *T-box* genes, another family of putative transcription factors differentially expressed in the forelimb and hindlimb buds<sup>13</sup>. Combinatorial action between genes may explain different functional requirements for the *HoxA* and *HoxD* genes in the forelimb and hindlimb. For example, homozygous deletion of both *HoxA-11* and *HoxD-11* results in almost complete loss of the zeugopod in the forelimb

but not in the hindlimb<sup>14</sup>, despite the fact that these genes have equivalent patterns of phase II expression in both limbs. A possible explanation for the observed differences between the appendages may be that there is expression of the paralogous gene *HoxC-11* during phase II in the hindlimb but not the forelimb, where it may act redundantly with *HoxD-11*.

**Learning to crawl: the fin-to-limb transition.** Some regions of vertebrate appendages are more variable than others<sup>4,15,16</sup>. The invention of flippers, wings and other specialized limbs often involved significant changes in the pattern of distal structures rather than proximal ones<sup>15</sup>. Two broad notions of the homology of distal structures have emerged over the past 130 years: one that sees digits as being unique to tetrapods<sup>17,18</sup> and another that sees antecedents of digital structure in the fins of sarcopterygian fish<sup>19,20</sup>. Both genetic and fossil data support the hypothesis that digits are evolutionary novelties<sup>21,22</sup> (Fig. 3).

The origin of digits is associated with the evolution of new temporal and spatial patterns of gene expression and regulation<sup>7,9,21</sup>. In extant tetrapods, the development of digits correlates with a reversal in the anteroposterior order of expression of *Hox* genes in phase II and phase III<sup>7,23</sup> (Figs 2, 3). Recent studies of teleosts



**Figure 1** Major innovations of vertebrate paired appendages. Basal chordates, such as *Amphioxus* (not shown), do not possess appendages homologous to those of vertebrates. Unpaired, median fins are the earliest known vertebrate appendages<sup>2</sup>. Paired appendages are first encountered in Ordovician and Silurian jawless fish as elongate fins that extend laterally along the body wall (for example, *Jamoytius*; **a**) or as paired pectoral fins (osteostracans; **b**). Other basal vertebrates (not shown) do not possess any paired appendages. Multiple sets of paired appendages are a derived characteristic of jawed fish (gnathostomes). In many gnathostomes, pectoral and pelvic fins have often evolved in parallel. This pattern of concerted evolution suggests that pectoral and pelvic appendages shared similar regulatory genes in early stages of gnathostome evolution. A fin axis (blue) is seen in the fins of many gnathostomes, and it is a primitive characteristic for sarcopterygian fish (for example, *Eusthenopteron*; **c**) and tetrapods (for example, *Ichthyostega*; **d**). Sarcopterygian fins are derived in having a zeugopod and an elaborate endoskeletal fin skeleton. Digits develop

within the distal portion of this extensive endoskeleton. The establishment of serially homologous appendages is proposed to result from gene cooption during the evolution of Paleozoic vertebrates. *HoxD* genes were probably not involved in the origins of body wall outgrowths in basal vertebrates because unpaired fins do not express these genes<sup>9</sup>. These *Hox* genes were initially involved in specifying regional identities along the primary body axis, particularly in caudal segments (**e**). One key step in the origin of jawed fish was the cooption of similar nested patterns of expression of *HoxD* genes in the development of both sets of paired appendages (**f**). This cooption may have happened in both appendages simultaneously, or *Hox* expression could have been initially present in a pelvic appendage and been coopted in the development of an existing pectoral outgrowth<sup>8</sup>. The reconstructions in **a** and **b** are modified from those in ref. 106, that in **c** from ref. 107, and that in **d** from ref. 108. The hind limb of *Ichthyostega* (**d**) is modified from ref. 5.

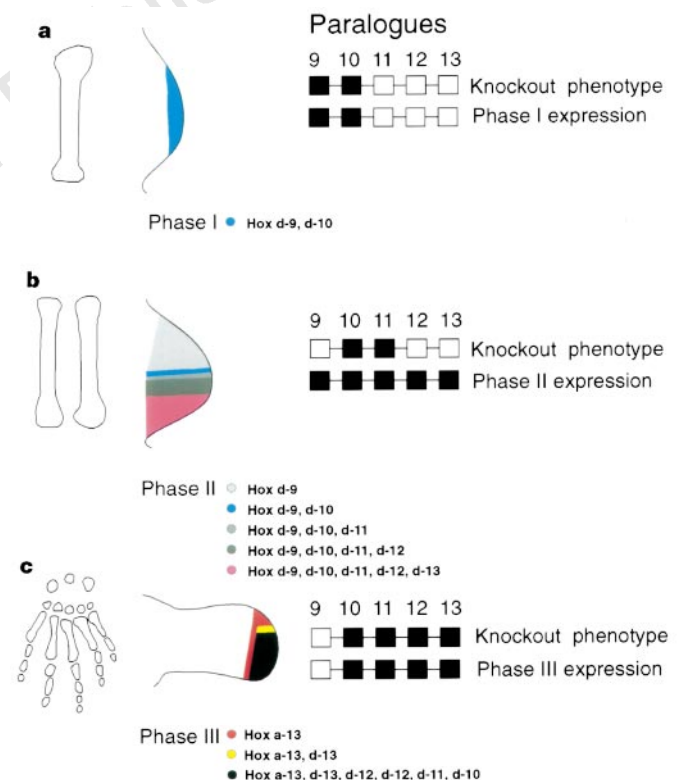
(zebrafish) have revealed patterns of *Hox* expression that are similar to patterns seen in proximal regions of tetrapod limbs<sup>9,21</sup>. Phase III *Hox* expression is not seen in the zebrafish and appears to be unique to the digital region of tetrapod limbs<sup>7,9</sup> (the expression of other sarcopterygians or more basal actinopterygians is not known). In addition, the different phases of *Hox* expression are not only discrete in tetrapod limbs, but they are regulated by separate *cis*-regulatory enhancer elements in each phase<sup>24–26</sup>. During phase II *Hox* expression in tetrapods, a complex set of enhancer elements is used within the regulatory region of each *Hox* gene of the cluster (as in the regulation of the *Hox* genes along the main body axis)<sup>25,26</sup>. However, regulation of all the *HoxD* genes in phase III depends upon a single enhancer upstream of the entire cluster<sup>24,26</sup>. The utilization of this distinct enhancer is consistent with the hypothesis that digits are evolutionary novelties because the development of the autopod is regulated differently from that of the rest of the limb.

The presence of phase III *Hox* expression in tetrapod limbs, and its absence in teleost fins, suggests that this pattern may be an apomorphy for tetrapods or a more inclusive group. In addition, the presence of a uniquely tetrapod enhancer for phase III *Hox* expression implies that this regulatory element is also more derived

relative to conserved phase I and II enhancers. The shift from phase II to phase III collinear expression involves multiple genes expressed at different times and in different regions of the limb. If these changes were genetically independent then they would have required the joint evolution of numerous regulatory elements. If only a single enhancer was involved, then this shift could have produced a change in the expression of multiple genes in a small number of evolutionary steps. Furthermore, the utilization of the same enhancer in forelimbs and hindlimbs provides a developmental explanation of the observation that fingers and toes arose simultaneously in the fossil record.

We propose that the temporal and spatial shift in the expression of *Hox* genes during limb development correlates with transformations inferred from the fossil record. Devonian fossils provide morphological links between structures in fins and limbs. Sarcopterygian fins are dominated by an axis of segmented endoskeletal elements that extends from proximal to distal<sup>2–4,27</sup> (Figs 1, 3). This axis is most similar to tetrapod limbs proximally, where the humerus, radius and ulna (femur, tibia and fibula) can readily be compared between taxa<sup>2–5,15,22</sup>. Embryological and palaeontological data suggest that the axis of fins was developmentally bent during

**Figure 2** Stylopod, zeugopod, and autopod: patterning the limb. The tetrapod limb consists of three distinct compartments: **a**, the stylopod (upper arm and thigh); **b**, zeugopod (lower arm and calf); and **c**, autopod (hand and foot). This subdivision of the limb is supported by phylogenetic comparison<sup>4</sup>, analysis of gene expression<sup>7</sup>, and experimental manipulation<sup>14,36–41,109–111</sup>. There is a broad correlation between the position of a compartment and its evolutionary history<sup>4</sup>. The stylopod (**a**) is the most ancient (possibly of Late Silurian origins) whereas the zeugopod and autopod are the most recent (being first encountered in Devonian sarcopterygians). This same order of appearance of the three limb segments is recapitulated during development. Early removal of the apical ectodermal ridge (AER) results in a limb with only a stylopod; the zeugopod and autopod are produced after successively later surgeries<sup>112,113</sup>. The *Abd-B*-related genes of the *HoxD* cluster are expressed in a complex, dynamic pattern encompassing at least three distinct<sup>7</sup>, independently regulated phases<sup>24,26</sup>. In the first phase (**a**; phase I), two of these genes (*HoxD*-9 and *HoxD*-10) are expressed across the entire limb bud<sup>7</sup>. This expression correlates with the time that the stylopod is specified<sup>7,12</sup>. Subsequently, a second phase of expression (**b**; phase II) is initiated in response to the secreted factor, *Sonic hedgehog*. Here, *Hox* genes are expressed in a nested set centred around the *Sonic*-expressing cells, with *HoxD*-13 being expressed in the most restricted domain, and *HoxD*-12 and *HoxD*-11 each encompassing a broader domain<sup>7</sup>. This pattern of expression coincides with the time of specification of the zeugopod and takes place in cells fated to form this segment. Finally, a third phase of expression (**c**; phase III) is initiated later during limb development, when these *Hox* genes are all expressed across the majority of the distal portion of the limb bud<sup>7</sup>. During this phase, the expression of *Hox* genes still appears to be a consequence of the *Sonic hedgehog* signal, but the relative responsiveness of the different genes has changed so that *HoxD*-13 now has the broadest expression domain and *HoxD*-12 and *HoxD*-11 are nested within it<sup>7</sup>. The phase III expression (**c**) patterns occur in the presumptive autopod at the time that segment is specified. The combination of the change in relative size of *Hox* expression domains with a phenomenon known as 'posterior prevalence' (the general rule that more 5' genes in the *Hox* cluster are phenotypically dominant) results in different *Hox* genes playing pre-eminent roles in different limb segments: for example, *HoxD*-9 during phase I in the stylopod (**a**), *HoxD*-11 during phase II in the zeugopod (**b**), and *HoxD*-13 during phase III in the autopod (**c**). The expression of the dominant *Hox* genes in each phase is essential for the formation of the bones in each segment, as seen in their knockout phenotypes. The knockout phenotype of a gene consists of alterations in the pattern and shape of skeletal elements. The location of these modifications depends on the position of the gene within the cluster. For example, mice engineered to be deficient in both *HoxD*-11 and in *HoxA*-11 (the paralogous gene of the *HoxA* cluster) form limbs that are essentially missing the zeugopod<sup>14</sup> (**b**). Phenotypes of *HoxD*-9-deficient mice, in contrast, are specific to the stylopod<sup>109</sup> (**a**), whereas *HoxD*-13 deficient mice primarily have defects in the autopod<sup>39</sup> (as do mice engineered to be deficient of *Hox A*-13 paralogues<sup>110,111</sup>) (**c**). Knockout data are derived from refs 14, 36–41 and 109–111.

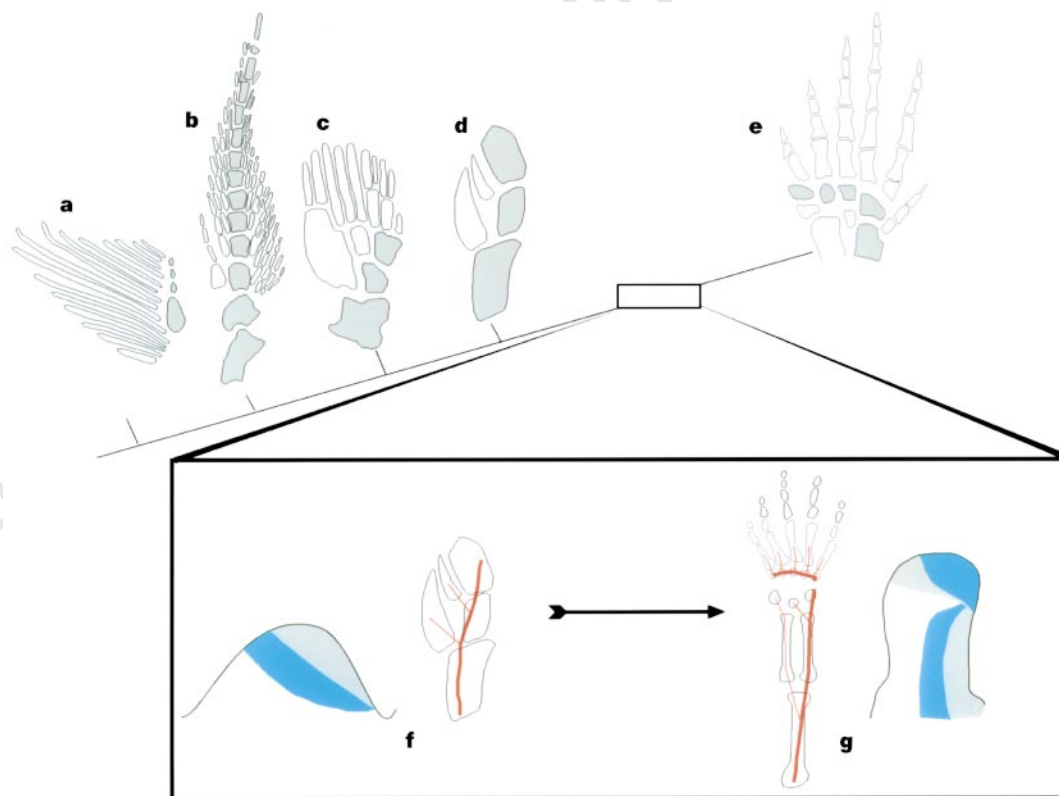


the origin of tetrapod limbs<sup>27</sup>. This scheme holds that there is a dramatic difference between the autopod and the zeugopod because the branching of the axis shifts from the anterior (preaxial) to the posterior (postaxial) compartment of the limb<sup>7,27</sup>. Proximal elements, such as the radius, project anteriorly from the axis, whereas distal elements, such as the digits, project from the posterior side of the axis. We propose that the reversal of morphological polarities in the appendages of Devonian vertebrates correlates with the reversal of *Hox* gene expression seen in phase III (Fig. 3f, g). As *Hox* expression in phase III is driven by a novel enhancer element, the axis was not bent *per se*; rather, a novel extension (with reversed morphological polarities) is considered to be added to it in the Late Devonian. This hypothesis is supported by a comparison of panderichthyid fins and tetrapod limbs<sup>20</sup> (Fig. 3). The fins of this sister-group of tetrapods (Fig. 3d, f) are highly reduced in comparison to other sarcopterygians (Figs 1e, 3b, c, c) and this reduction is most prominent distally. No potential homologues of digits, wrist or ankle bones are preserved in these fish<sup>22</sup>.

Are digits, or their functional equivalents, unique to tetrapods? Fins of rhizodontid fish have stunning similarities to tetrapod limbs<sup>20</sup> (Fig. 3). The fins of these Devonian fish contain up to eight endoskeletal radials that project distally; in *Sauripteris* (Fig. 3c), six of these rods terminate at the same proximodistal level. Either

these radials are directly homologous to the six to eight digits of Devonian tetrapods (a hypothesis not supported by phylogenetic inference) or they are functional analogues. In either case, these rods reflect a site at which morphological polarities are reversed (for example, the radials of *Sauripteris* branch postaxially<sup>20</sup>), suggesting that phase III *Hox* expression may have arisen in this clade. Several different lineages of Devonian sarcopterygians appear to have evolved the same morphological solution to life in shallow freshwater environments. The tetrapod clade evolved true digits, whereas rhizodontids developed functional analogues. In both cases, the genetic shifts may well have been similar.

**Adaptive diversification: how many fingers?** Digit reduction is a dominant theme of tetrapod limb evolution; deviations from a pentadactyl pattern virtually always involve the loss of fingers or toes<sup>5,28</sup>. Polydactylous hands and feet have almost never been fixed in phylogeny, despite the presence of polydactylous variants within populations (individuals of many species including cats, dogs, mice, chickens and humans carry mutations that cause the formation of extra digits). This paradox can be explained in terms of evolutionary constraints by postulating a genetic limitation to digital evolution. One approach holds that the genetic mechanisms that determine the number of digits are distinct from those that regulate morphology, and that there are currently only five discrete genetic programs for



**Figure 3** The origin of digits. A fin axis (grey) is present in the fins of chondrichthyans (a, *Cladoselache*), basal actinopterygians (not shown), and sarcopterygians (b–e). In sarcopterygians such as *Neoceratodus* (b), *Sauripteris* (c), *Panderichthyes* (d) and *Tulerpeton* (e), a single element (stylopod) articulates with the pectoral or pelvic girdle; all other proximal bones of the fin have been lost. The autopod is considered to be a synapomorphy of tetrapods because their nearest relatives do not have any apparent homologues of digits. Box: we propose that the origin of digits correlates with a novel pattern of *Hox* expression. f, The axis (red) of the fin of *Panderichthyes* is short and radials branch preaxially. Patterns of phase II *Hox* expression (f; *Hox D-11* in light blue, *Hox D-13* in darker blue) were most likely already in place in sarcopterygian fins. g, The proximal portion of the axis (red) of tetrapods compares with the entire axis of *Panderichthyes*. Phase III *Hox* expression reflects a reversal of the nested

domains of expression of *HoxD-11* (g, light blue) and *HoxD-13* (g, dark blue) during the time when the autopod is specified. This reversal in the polarity of *Hox* expression is considered to be correlated with the origin of the autopod; digits differ from more proximal structures in that they lie on the postaxial side of the axis. Because this hypothesis relies on genetic comparisons between phylogenetically disparate taxa (for example, teleosts and tetrapods), the shift from phase II to phase III collinear expression may have evolved in more basal sarcopterygians. Analogues of digits are seen in the fins of other sarcopterygian fish (c, *Sauripteris*); the expanded endoskeletons of rhizodontids include as many as eight branched preaxial radials. Different lineages of sarcopterygians appear to be inventing similar solutions to life in shallow freshwater ecosystems. The reconstruction of *Tulerpeton* (e) is modified from ref. 114.



Are regularities of digital evolution the product of developmental constraints upon variation? Knockouts of different *Hox* genes (*HoxD-11*, *HoxD-12*, *HoxD-13*, *HoxA-13*) lead to changes in the shape and number of bones in affected mice and these different genes often have overlapping effects<sup>14,36–41</sup>. One common result is the stability of the internal digits (IV, III) in knockouts of single genes or combinations of genes<sup>41</sup>. The parallels between the expectations of Morse's law and the results of experimental manipulation suggest that trends of digital evolution may have a developmental basis. The morphological effects of different gene knockouts may reflect the sequence of digital formation: the first digits to be affected are typically the last to form in development<sup>37</sup>. Although evolutionary patterns of digital reduction are unlikely to involve coding mutations of *Hox* genes, limb reduction may involve changes

Morphological studies of Cambrian lobopodans have investigated the series of innovations that led to the basal arthropod design. These novelties include: the evolution of external segmentation; sclerotization; and, most important to our discussion, the origin of jointed, biramous appendages. Whereas the taxonomic relationships between Cambrian forms and extant arthropods are uncertain, Cambrian taxa with different types of 'arthropodization' and

The diagram illustrates a phylogenetic tree of Chelicerata and other groups, with morphological characters and gene expression patterns mapped onto the branches. The tree is rooted at the bottom with Onychophora and branches upwards. The groups shown are Onychophora, armoured lobopods, Opabinia, Anomalocaris, chelicerates, myriapods, crustacea, and insects. The morphological characters mapped onto the tree are: heteronomous annulation (Onychophora), lateral lobes, segmentation, tail fan (armoured lobopods), appendage sclerotization; biramous jointed limbs (Opabinia), cuticular sclerotization (Anomalocaris), uniramious limbs (chelicerates, myriapods, crustacea, insects), and uniramious limbs, mandible reduced (insects). The gene expression patterns mapped onto the tree are: Sor (insects), Anip (insects), Ubx (insects), abdA (insects), and AbdE (insects). The gene expression patterns for Anomalocaris and Opabinia are also shown, with Anip, Ubx, abdA, and AbdE expressed in the appendages.

limb morphology can be identified<sup>43</sup>. For example, the Cambrian *Aysheaia* and the Recent *Peripatus*, possess simple, unjointed uniramous limbs (Fig. 4), but lack the annulation and armour typical of more derived lobopodans such as *Hallucigenia*<sup>43</sup> (Fig. 4). Other lobopodans, such as *Opabinia*, are externally segmented and possess lateral lobes above the ventral lobopods<sup>43</sup> (Figs 4, 5b). Fusion between the gill-like lateral lobes and ventral lobopodia may have given rise to the biramous limb<sup>43,45–47</sup> (Fig. 4c). In forms such as *Anomalocaris*, this fusion was accompanied by limb segmentation and sclerotization (Figs 4, 5). Full cuticular sclerotization, then, arose in primitive arthropods.

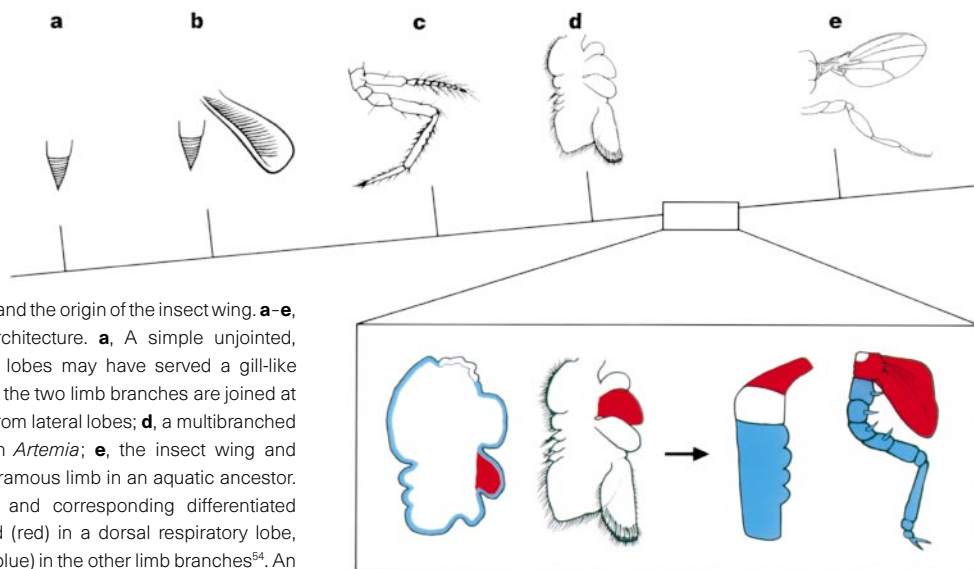
**Serial homology and adaptive diversification.** The adaptive radiation of primitive arthropod limbs entailed considerable changes in their number, pattern and function. Two trends are evident. First, in arthropods in general and certain lineages in particular, there have been increases in the number of different limb types. For example, advanced lobopods had perhaps four types of appendages (frontal, jaw, trunk and tail fan) and only one type of trunk appendage, whereas insects possess up to ten distinguishable limb-derived appendages and four or five different trunk appendages (Fig. 4). A second trend has been the dramatic diversification of homologous appendages—not just in size or morphological detail. Major changes in limb organization have evolved, such as the evolution of unbranched walking legs in insects and changes in mandible architecture in myriapods, crustacea and insects (Fig. 4).

The developmental genetics of limb formation and identity in Recent arthropods may help to shed light on the diversification of arthropod limbs. *Hox* genes have played a different role in arthropod evolution from that seen in tetrapods. Studies of *Drosophila melanogaster* have revealed that each type of appendage is typically specified by a single or a pair of *Hox* genes acting in the individual body segment that gives rise to a particular appendage<sup>48</sup>. For example, the *Antennapedia* gene acts in all three pairs of walking legs but the distinct morphology of the first, second and third pair of walking legs is determined by the *Sex combs reduced*, *Antennapedia*, and *Ultrabithorax* *Hox* genes<sup>49</sup>, respectively. In the antenna, no *Hox* gene is active. If *Hox* gene function is lost or ectopically activated in individual segments, the identity of the corresponding appendage is transformed. Thus, loss of *Antennapedia* transforms second leg to antennal structures<sup>50</sup> and expression of *Antennapedia* in the antenna

transforms it to a leg<sup>51</sup>. Importantly, loss of all *Hox* gene functions in insects results in a dead embryo bearing antennae on all segments<sup>52</sup>. This demonstrates that the potential to form a limb exists in all segments, but the type of limb formed is determined by individual *Hox* genes.

The specification of different limb types by different *Hox* genes or combinations of *Hox* genes differs from the nested pattern of *Hox* genes expressed in the limbs of vertebrates and has important implications for the pattern of morphological evolution in the arthropods. Different, but serially homologous, arthropod limbs are distinguished by the action of different *Hox* genes that modify the interpretation of a common set of positional signals. Thus, the increase in appendage diversity between lobopods, primitive crustacea and insects must have involved the diversification of *Hox* gene regulation and function in the arthropod trunk. Comparative studies of *Hox* gene expression between crustacea (for example, branchiopods) and insects support this notion. In the branchiopod thorax, the expression of the *Antp*, *Ubx* and *abd-A* *Hox* genes is coincident and the morphology of thoracic limbs is uniform<sup>53</sup>. In insects, these same *Hox* genes differentiate the middle thorax, posterior thorax and anterior abdomen (Fig. 5). The adult hexapod abdomen is legless and this is due to the direct repression of limb formation by products of the *Ubx* and *abd-A* *Hox* genes<sup>54</sup>. These gene products do not repress limb formation in branchiopods<sup>55</sup>.

The diversity of the architecture of putatively homologous appendages has fuelled many debates about arthropod relationships. For example, walking legs and mandibles can differ so much between taxa that it has been suggested that the arthropods had multiple ancestors (that is, they are polyphyletic)<sup>56</sup>, but phylogenetic<sup>57,58</sup> studies have refuted this. In addition, developmental studies suggest that different limb architectures arise through modifications of a common genetic program. For example, the *Distal-less* (*Dll*) gene controls the development of the distal portion of *Drosophila* limbs<sup>59</sup> and is expressed in the distal domains of limbs in all arthropods studied so far<sup>50,60,61</sup>. However, *Dll* is not expressed in insect or developing adult crustacean mandibles<sup>55</sup> but is expressed in myriapod mandibles<sup>62</sup>. These data agree with fossil evidence suggesting that crustacean and insect mandibles were reduced from the primitive whole-limb mandible by truncation of the mandibular proximodistal axis.



**Figure 5** The evolution of the arthropod limb and the origin of the insect wing. **a–e**, Some of the major transitions in limb architecture. **a**, A simple unjointed, annulated lobopodium; **b**, separate lateral lobes may have served a gill-like function; **c**, a jointed biramous limb in which the two limb branches are joined at the base, an upper branch is often derived from lateral lobes; **d**, a multibranching limb found in the branchiopod crustacean *Artemia*; **e**, the insect wing and uniramous leg appear to derive from a polyramous limb in an aquatic ancestor. Box: left, a developing polyramous limb and corresponding differentiated structure. The *apterous* gene is expressed (red) in a dorsal respiratory lobe, whereas the *Distal-less* gene is expressed (blue) in the other limb branches<sup>54</sup>. An ancestor–descendent relationship between the limbs in the box is not implied. Right, separation of the dorsal respiratory lobe from the ventral limb primordium in a primitive pterygote such as a Paleodictoptera nymph. The proto-wing at this stage was probably a gill-like structure on all trunk segments and still attached to the base of the limb. The *apterous* and *Distal-less* genes play critical roles in wing and leg formation in *Drosophila*.

The branching of arthropod limbs has been very important to their functional evolution. Different limb branches can be specialized for respiration, locomotion and a variety of other functions. Chelicerates, trilobites and aquatic crustacea have biramous or polyramous limbs and first appear in the Cambrian, whereas the terrestrial myriapods, insects and crustacea have unbranched (uniramous) limbs and appear much later in the Silurian<sup>63</sup> and Devonian<sup>64</sup>, respectively. It has been argued that the ancestral arthropod was biramous<sup>43,45–47</sup>. Comparisons of *Dll* expression and limb outgrowth in various types of crustaceans and insects reveal that all limb types arise from the same relative anteroposterior position within the body segment but differ in their dorsoventral branch points<sup>55</sup>. This suggests that uniramy, biramy and polyramy are the products of shifts in signals along the dorsoventral axis of the body wall (or appendage) and that additions or reductions in branch number may evolve readily. This flexibility was crucial to the later evolution of perhaps the most significant invention by any arthropod—wings. **Learning to fly: the leg-to-wing transition.** Early in the Devonian, before tetrapods arose, one major animal group had already invaded land—the insects. The subsequent evolution of flight presaged the enormous radiation of insects: these taxa now comprise more than two-thirds of all known animal species. The evolution of insects from an as-yet uncertain arthropod ancestor and the emergence of winged (pterygote) forms involved major transitions in limb architecture and function.

Few evolutionary mysteries have inspired more theories than the origin of insect wings. It is not certain when wings first arose because the Early Devonian insect fossil record is scanty<sup>64</sup>. All of the major pterygote groups appear by the Carboniferous (362–290 Myr) and are assumed to have arisen earlier. Fundamentally, discussion on the origin of wings has focused on whether they are novelties or whether they are modified versions of ancestral structures. If wings were derived from existing structures, what were their anatomical origins and initial functions?

One of the longest-held models is the ‘paranotal’ theory, which holds that wings are novelties derived from hypothetical rigid extensions of the body wall of a terrestrial ancestor<sup>65</sup>. A second hypothesis, the ‘limb-exite’ model proposes that insect wings evolved in a series of transitions beginning with the polyramous exite-bearing legs of an aquatic pterygote ancestor<sup>42,66</sup>. According to this model, proximal limb elements were modified to flap-like structures and adapted to spiracular or movable gill covers to facilitate respiration. These sack-like pro-wings were found on all thoracic and abdominal segments and became stronger as ancestral pterygote nymphs (similar to Recent mayfly larvae) used them presumably for propulsion. As insect lifestyles became more amphibiotic, some insects might have used a rudimentary proto-wing for surface skimming, as seen in extant stoneflies<sup>67</sup>. Finally, wings acquired the mechanical strength and flexibility (with corrugation and veins) and the supporting musculature to support active flight.

Recent studies of the developmental regulatory mechanisms controlling wing formation and number in *Drosophila* suggest a close ontogenetic and evolutionary relationship between legs and wings. For example, in *Drosophila*<sup>68</sup> and other Diptera, the wing arises in close association with the leg. Cells that give rise to the wing field in *Drosophila* actually migrate out of the developing ventral limb field<sup>68</sup>. There are important regulatory differences that distinguish the development of the sheet-like wing from the tubular structure of the leg. One key difference is that adult and developing wings are divided into discrete dorsal and ventral compartments whereas legs are not. In *Drosophila*, the definition of dorsal versus ventral cell fates is orchestrated by the *apterous* gene which is expressed only in dorsal cells and is necessary for wing but not leg formation<sup>69</sup>. Clearly, *apterous*, which regulates several crucial downstream signalling components, and is involved in a conserved dorsal pattern of expression in insect wings<sup>70</sup>, was coopted into a distinct role in dorsoventral patterning at some stage of wing evolution.

*apterous* primitively could have specified a dorsal compartment (or branch) of an ancestral polyramous appendage or an evolving proto-wing. Remarkably, this is exactly what has been found in a recent study<sup>71</sup>. The dorsal branch of a branchiopod crustacean respiratory epipodite specifically expresses the *apterous* gene and one other developmental marker of the insect wing field (Fig. 5d). This suggests that Recent wings evolved from the respiratory lobe of an ancestral polyramous limb, probably first appearing in the immature aquatic stages as gill-like structures, such as those found on all trunk segments of extinct Paleodictyoptera or extant mayfly larvae (Fig. 5d). Wings subsequently emerged as adult appendages and acquired greater strength and flexibility for sustained flight (Fig. 5e).

Interestingly, if the respiratory epipodite origin of insect wings is correct, then wings may have an even deeper origin, not just in the aquatic ancestor of pterygotes, but in lobopodans. The origin of the biramous limb has been postulated to involve the fusion of the gill-like structure of lateral lobes (such as those in *Opabinia*) with the ventral lobopod<sup>43,45–47</sup> (Fig. 5c). If this is true, then the wing may indeed be derived from lateral lobes, not of a terrestrial insect as once thought, but of a much more distant lobopodan ancestor.

### Deep homology and origin of appendages

It is clear from the fossil record that chordates and arthropods diverged at least by the Cambrian. The appendages of these two groups are not homologous because phylogenetically intermediate taxa (particularly basal chordates) do not possess comparable structures. The most surprising discovery of recent molecular studies, however, is that much of the genetic machinery that patterns the appendages of arthropods, vertebrates and other phyla is similar. These findings suggest that the common ancestor of many animal phyla could have had body-wall outgrowths that were organized by elements of the regulatory systems found in extant appendages. We now describe these similarities and use them to consider the origin of animal limbs.

In *Drosophila*, the anteroposterior (AP) axis of the leg or wing imaginal disc (the larval precursor to the adult appendages) is divided into two compartments (reviewed in ref. 49). The posterior half of the disc expresses the gene *hedgehog*<sup>72,73</sup>, which encodes the key signal that initiates AP patterning (Fig. 6). In response to *hedgehog*, a thin layer of cells running along the border of the anterior and posterior compartments is induced to produce another secreted protein encoded by the gene *decapentaplegic* (*dpp*)<sup>74</sup>. *dpp*, in turn, is a long-range signal providing positional information, and hence differential AP fates, to cells in both compartments<sup>75–79</sup>. Misexpression of either *hedgehog* or *dpp* in the anterior of the disc results in AP mirror-image duplications of limb structures<sup>74</sup>.

The AP axis of the vertebrate limb is set up in a very similar manner (Fig. 6). A key organizing signal is *Sonic hedgehog* (*Shh*), one of the three direct homologues of the *Drosophila* gene *hedgehog*<sup>80–83</sup>. Like *hedgehog*, *Shh* is localized posteriorly in the limb bud. Misexpression of *Shh* anteriorly causes AP mirror-image duplications analogous to those caused by *hedgehog* misexpression in the fly imaginal disc<sup>74,83</sup>. In addition, *Bmp-2*, one of the two vertebrate homologues of the arthropod *dpp* signalling protein, is expressed in the limb bud in response to *Shh*<sup>83</sup>. Unlike *Drosophila dpp*, *Bmp-2* does not have the ability to cause full limb duplications. However, it clearly functions as a secondary signal in the *Shh* pathway, polarizing the overlying ectoderm<sup>84</sup>.

Signals organizing proximodistal (PD) outgrowth in vertebrate and insect appendages also operate similarly. In the insect imaginal wing disc, PD outgrowth is organized by a specialized set of cells running the length of the dorsoventral (DV) border, the ‘wing margin’ (Fig. 6). The dorsal compartment of the wing is characterized by the expression of the transcription factor *apterous*. *apterous* specifies dorsal-specific cell fate<sup>69</sup> and controls the expression of a secreted protein called *fringe*<sup>85</sup>. The interface between cells expressing



and cells not expressing *fringe* becomes the wing edge or margin<sup>85</sup>. A key downstream effector of *fringe* activity is encoded by *Serrate*<sup>86</sup>. In response to *fringe*, *Serrate* is induced, leading to the activation of several downstream effector genes<sup>87–90</sup> and the production of a signal at the margin which organizes the growth of the wing blade<sup>87–89</sup>.

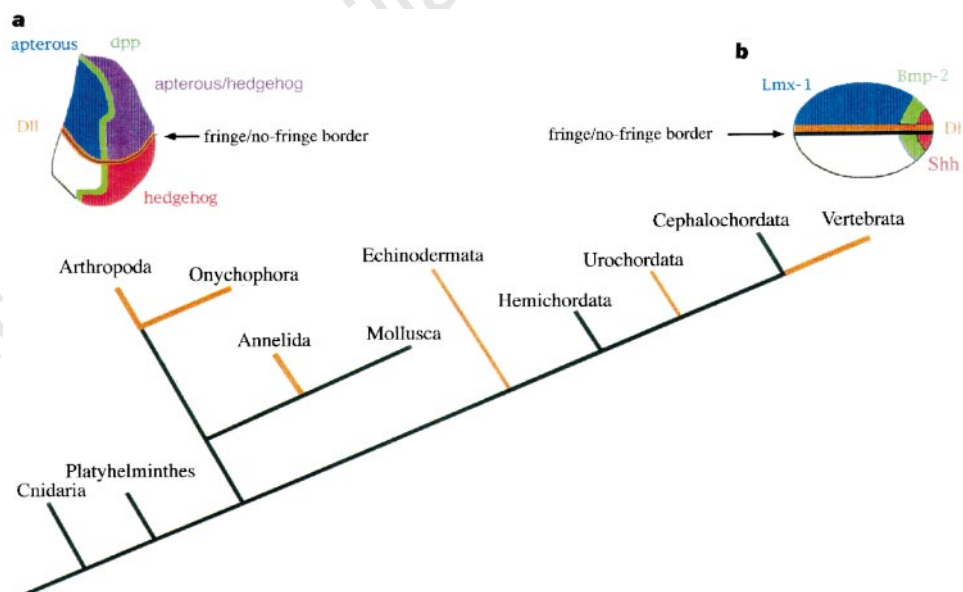
Unexpectedly, outgrowth of vertebrate limbs appears to be established by a very similar genetic cascade. Outgrowth of the limb bud is driven by signals from a specialized ectodermal structure, the apical ectodermal ridge<sup>91</sup> (AER), which, like the wing margin, runs along the DV border of the limb. Remarkably, a vertebrate homologue of *fringe*, called *Radical-fringe*, is expressed in the dorsal half of the limb ectoderm prior to formation of the AER<sup>92</sup>. At the border between cells expressing *Radical-fringe* and cells not expressing *Radical-fringe*, a homologue of *Serrate*, *Ser-2*, is induced and the AER forms. A *Radical-fringe* boundary is required to form the AER and ectopic *radical fringe* can induce an additional AER on the ventral surface<sup>93</sup>.

There are also parallels between the regulation of the DV axis in vertebrate and arthropod appendages (Fig. 6). Genes specifying DV polarity in both groups have been identified. In *Drosophila*, the early ventral expression of the gene *wingless*, a member of the *Wnt* family of secreted factors, is necessary for the proper DV patterning of the wing<sup>94,95</sup>. Subsequently, the expression of the transcription factor *apterous* defines the dorsal compartment and specifies dorsal cell fates<sup>69</sup>. In the vertebrate limb, the early expression of a different *Wnt* family member is also required for DV patterning. *Wnt7a* is specifically expressed throughout the dorsal ectoderm and is necessary and sufficient for many aspects of dorsal patterning<sup>96–98</sup>.

*Wnt7a* acts by inducing mesodermal expression of *Lmx-1*<sup>97,98</sup>. Like *apterous*, *Lmx-1* is a related member of the LIM-homeodomain family of transcription factors. As with *apterous* (*Drosophila*), *Lmx-1* (vertebrates) expression defines a dorsal compartment, being expressed early throughout the dorsal half of the limb bud, and it is sufficient to convey dorsal cell fate<sup>97,98</sup>.

The simplest phylogenetic implication to draw from these comparisons is that individual genes that are expressed in the three orthogonal axes are more ancient than either insect or vertebrate limbs (Fig. 6). Indeed, several of the regulatory systems seen in arthropod and vertebrate limbs are also involved in the development of other organs in a variety of taxa. The phylogenetic distribution of regulatory circuits and morphological structures presents two major interpretations: either similar genetic circuits were convergently recruited to make the limbs of different taxa or a set of these signalling and regulatory systems are ancient and patterned a structure in the common ancestor of protostomes and deuterostomes<sup>99</sup>.

The first model holds that genes and/or genetic circuits were convergently recruited for limb development during the evolution of vertebrates and arthropods. These genes would not be involved in appendage development in the common ancestor of vertebrates and arthropods; each gene or circuit was involved in other developmental events. This notion would require the parallel cooption of members of similar gene families, acting along different developmental axes to pattern an outgrowth of the body wall in at least two taxa. The evolution of limbs in each group would, then, have involved the convergent recruitment of numerous genes to define



**Figure 6** A cladogram of selected metazoans shows the distribution of major genes involved with appendage development. Homologous signals are deployed in similar locations in the limb primordia of arthropods (**a**, *Drosophila*) and vertebrates (**b**, chick). Equivalent orientations (dorsal up, anterior left) of a chick left wing bud and a *Drosophila* left wing imaginal disc are shown. *Sonic hedgehog* (red) in the chick, and its homologue *hedgehog* (red and purple, where coexpressed with *apterous*) in the fly are produced in a posterior domain. These factors induce the expression of secondary patterning signals in the appendages: overlapping expression of *Bmp-2* (green) in the chick and adjacent expression of *dpp* (green), the *Bmp-2*-homologue, in the fly. Dorsal cell fates are specified in both systems by LIM homeodomain transcription factors expressed throughout the dorsal half of the appendage primordia: *Lmx-1* (blue) in the chick, and *apterous* (blue and purple, where coexpressed with *hedgehog*) in the fly. The outgrowth of both appendages is driven by a specialized group of cells (the AER in the chick and the wing margin in the fly) running along the anteroposterior axis at the junction of the dorsal and ventral compartments (yellow). These key groups

of cells are specified in both the chick and the fly by the border between dorsal cells expressing the gene *fringe* and ventral cells not expressing *fringe*. For simplicity in viewing, conserved genes in signal transduction (such as the *hedgehog* receptor, *patched*) and other parallels between the two systems (such as the expression of *Wnt* genes) are not shown. *Distal-less* homologues (*Dll* in *Drosophila* and *Dlx* in the chick) are drawn in orange. *Distal-less* orthologues are expressed in a wide variety of animal appendages, including the lobopods of onychophorans, the tube feet of echinoderm, and the wings of birds and flies (orange). The limbs of these taxa are not homologous as appendages because phylogenetically intermediate groups do not possess comparable structures. This suggests at least two phylogenetic possibilities: either similar genetic circuits were convergently recruited to make the limbs of different taxa, or these signalling and regulatory systems are ancient and patterned a different structure (presumably another type of outgrowth) in the common ancestor of protostomes and deuterostomes<sup>99</sup>.

similar developmental axes. If confirmed, this hypothesis would provide a stunning case of convergent evolution.

The second model is that some of these genes or circuits were components of an ancestral genetic regulatory system that was used to pattern a structure in the common ancestor of vertebrates and arthropods. This ancestral structure need not have been homologous to arthropod or vertebrate limbs; the regulatory system could have originally patterned any one of a number of outgrowths of the body wall in a primitive bilaterian for example. The genes themselves were initially involved in other developmental events; the key step in animal limb evolution was the establishment of an integrated genetic system to promote and pattern the development of certain outgrowths. Once established, this system provided the genetic and developmental foundation for the evolution of structures as diverse as wings, fins, antennae and lobopodia.

The evaluation of these two alternative models requires consideration of several factors that could affect the identity and deployment of appendage-patterning genes. First, given the independent histories spanning more than 500 Myr of the lineages being compared, one should expect regulatory differences among patterning systems. Even the insect wing and leg, which are both probably derived from an ancestral polyramous appendage, have acquired different patterning mechanisms. Because individual regulatory components appear to have been gained, lost and modified during insect wing and leg evolution there is no *a priori* reason to expect that either structure would be more genetically similar to vertebrate limbs. Second, one member of a gene family may substitute for another during normal development. Indeed, some of the genes that are deployed similarly in arthropods and vertebrates are not strictly orthologous. Such substitution could be the product of either convergent evolution or descent with modification (by substitution) among redundant genes. Third, it could be argued that the presumed inversion of the DV body axis during deuterostome evolution would imply a corresponding inversion of appendage DV axis patterning signals, an expectation that is not met by the observed expression of *apterous*- and *fringe*-related genes in vertebrate limb buds. A parallel in reversal of the DV axes of the body and the limbs would be expected if vertebrate and insect limbs were structurally homologous to a common ancestral appendage which predated the reversal. However, we know that the modern vertebrate limb evolved after the DV inversion of the body axis. Thus, a DV inversion in the patterning of the limbs would not necessarily be predicted, whether the axis patterning genes were independently coopted for appendage formation according to the first model, or whether they were coopted as a unit, but regulated independently of the body axis, according to the second model.

One can argue many ways from the comparison of only two taxa: the alternative phylogenetic hypotheses need to be tested by additional comparative data. Evidence in support of an ancient common mechanism for the formation of outgrowths of the body wall comes from phylogenetic comparison of the expression of the transcription factor *Distal-less* (*Dll*)<sup>100</sup> (Fig. 6). *Dll* is expressed at the distal end of growing insect limbs<sup>55,60,61</sup>, and is essential for appendage outgrowth<sup>59</sup>. *Dll* orthologues are expressed in the distal portion (AER) of the embryonic limb buds of vertebrates, the ampullae and siphons of tunicates, the tubefeet of echinoderms, the parapodia of annelids, as well as Onychophoran lobopodia<sup>100</sup>. The expression of *Dll*-related genes could represent convergent utilization of the gene. However, the fact that out of the hundreds of transcription factors that potentially could have been used, *Dll* is expressed in the distal portions of appendages in six coelomate phyla makes it more likely that *Dll* was already involved in regulating body wall outgrowth in a common ancestor of these taxa (Fig. 6). The additional parallels between vertebrate and arthropod limbs suggest that this ancestral outgrowth may have also been patterned along the three orthogonal axes.

If a conserved outgrowth patterning system was available for co-

option in the evolution of vertebrate limbs, then it must have been used in patterning non-limb outgrowths in basal taxa. Genetic studies provide an example of at least one secondary outgrowth patterned along these axes that predated the evolution of vertebrate limbs: the branchial arches. As the branchial arches grow out from the cranial region of the chick embryo, they express important components of the limb patterning system in similar developmental regions<sup>101,102</sup>. Like arthropod and vertebrate limbs, the branchial arches contain localized, posterior expression of a *hedgehog* gene, in this case *Shh*<sup>101,102</sup>. Furthermore, *Shh* is coexpressed with *Bmp* posteriorly<sup>101</sup>. Yet more similarities lie in the DV and AP axes: *fringe*-expressing cells are initially confined to the dorsal ectoderm and later are restricted to distal regions of the outgrowth where *Distal-less* orthologues are also expressed<sup>93</sup>. The ectopic deployment and modification of an existing patterning program, such as that of the branchial arches, may have given rise to the predecessors of vertebrate appendages.

Determination of whether two structures are homologous depends on the hierarchical level at which they are compared<sup>103–105</sup>. For example, bird wings and bat wings are analogous as wings, having evolved independently for flight in each lineage. However, at a deeper hierarchical level that includes all tetrapods, they are homologous as forelimbs, being derived from a corresponding appendage of a common ancestor. Similarly, we suggest that whereas vertebrate and insect wings are analogous as appendages, the genetic mechanisms that pattern them may be homologous at a level including most protostomes and deuterostomes. Furthermore, we propose that the regulatory systems that pattern extant arthropod and vertebrate appendages patterned an ancestral outgrowth and that these circuits were later modified during the evolution of different types of animal appendages. Animal limbs would be, in a sense, developmental 'paralogues' of one another; modification and redeployment of this ancient genetic system in different contexts produced the variety of appendages seen in Recent and fossil animals. □

Neil Shubin is in the Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; Cliff Tabin is in the Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA; and Sean Carroll is at the HHMI and Laboratory of Molecular Biology, University of Wisconsin, 1525 Linden Drive, Madison, Wisconsin 53706, USA

- Müller, G. B. & Wagner, G. P. Novelty in evolution: Restructuring the concept. *Ann. Rev. Ecol. Syst.* **22**, 229–256 (1991).
- Coates, M. I. The origin of vertebrate limbs. *Development* (suppl.) 169–180 (1994).
- Coates, M. I. Fish fins or tetrapod limbs—a simple twist of fate? *Curr. Biol.* **5**, 844–848 (1995).
- Shubin, N. The evolution of paired fins and the origin of tetrapod limbs. *Evol. Biol.* **28**, 39–85 (1995).
- Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal tetrapod interrelationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb.* **87**, 363–421 (1996).
- Johnson, R. & Tabin, C. The long and short of *hedgehog* signaling. *Cell* **81**, 313–316 (1995).
- Nelson, C. E. *et al.* Analysis of *Hox* gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).
- Tabin, C. J. & Laufer, E. *Hox* genes and serial homology. *Nature* **361**, 692–693 (1993).
- Sordino, P., van der Hoeven, F. & Duboule, D. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–681 (1995).
- Kessel, M. & Gruss, P. Murine developmental control genes. *Science* **249**, 374–379 (1990).
- Mackem, S., Ranson, M. & Mahon, K. Limb-type differences in expression domains of certain chick *Hox-4* genes and relationship to pattern modification for flight. *Prog. Clin. Biol. Res.* **383**, A, 21–30 (1993).
- Peterson, R. J., Papenbrock, T., Davada, M. M. & Awgulewitsch, A. The murine *Hoxc* cluster contains five neighboring *abdB*-related *Hox* genes that show unique spatially coordinated expression in posterior embryonic subregions. *Mech. Dev.* **47**, 253–260 (1994).
- Gibson-Brown, J. J. *et al.* Evidence of a role for *T-box* genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. *Mech. Dev.* **56**, 93–101 (1996).
- Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S. & Capocchi, M. R. Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. *Nature* **375**, 791–795 (1995).
- Vorobyeva, E. & Hinchliffe, J. R. From fins to limbs. *Evol. Biol.* **29**, 263–311 (1996).
- Hinchliffe, J. R. & Johnson, D. R. *The Development of the Vertebrate Limb* (Clarendon, Oxford, 1980).
- Holmgren, N. On the origin of the tetrapod limb. *Acta Zoologica* **14**, 185–295 (1933).
- Holmgren, N. Contribution to the question of the origin of the tetrapod limb. *Acta Zoologica* **20**, 89–124 (1939).
- Watson, D. M. S. On the primitive tetrapod limb. *Anat. Anzeiger* **44**, 24–27 (1913).
- Gregory, W. K. & Raven, H. C. Studies on the origin and early evolution of paired fins and limbs. *Ann. N. Y. Acad. Sci.* **42**, 273–360 (1941).
- Sordino, P. & Duboule, D. A molecular approach to the evolution of vertebrate paired appendages. *Trends Ecol. Evol.* **11**, 114–119 (1996).
- Ahlberg, P. E. & Milner, A. R. The origin and early diversification of tetrapods. *Nature* **368**, 507–512 (1994).
- Yokouchi, Y. *et al.* Homeobox gene expression correlated with the bifurcation process of limb cartilage development. *Nature* **353**, 443–445 (1991).

24. Gerard, M., Duboule, D. & Zakany, J. C. Cooperation of regulatory elements involved in the activation of the *Hoxd-11* gene. *Compt. R. Acad. Sci. III* **316**, 985–994 (1993).
25. Beckers, J., Gerard, M. & Duboule, D. Transgenic analysis of a potential *Hoxd-11* limb regulatory element present in tetrapods and fish. *Dev. Biol.* **180**, 543–553 (1996).
26. van der Hoeven, F., Zakany, J. & Duboule, D. Gene transpositions in the *HoxD* complex reveal a hierarchy of regulatory controls. *Cell* **85**, 1025–1035 (1996).
27. Shubin, N. & Alberch, P. A morphogenetic approach to the origin and basic organization of the tetrapod limb. *Evol. Biol.* **20**, 318–390 (1986).
28. Tabin, C. J. Why we have (only) five fingers per hand: *hox* genes and the evolution of paired limbs. *Development* **116**, 289–296 (1992).
29. Holder, N. Developmental constraints and the evolution of vertebrate digit patterns. *J. Theor. Biol.* **104**, 451–471 (1983).
30. Morse, E. On the tarsus and carpus of birds. *Ann. Lyc. Nat. Hist.* **10**, 141–158 (1872).
31. Shubin, N., Crawford, A. & Wake, D. Morphological variation in the limbs of *Taricha granulosa* (Caudata: Salamandridae): Evolutionary and phylogenetic implications. *Evolution* **49**, 874–884 (1995).
32. Greer, A. Limb reduction in the Scincid lizard genus *Lerista*. 2. Variation in the bone complements of the front and rear limbs and the number of postsacral vertebrae. *J. Herpetol.* **24**, 142–150 (1980).
33. Lande, R. Evolutionary mechanisms of limb loss in tetrapods. *Evolution* **32**, 73–92 (1978).
34. Gauthier, J. Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).
35. MacFadden, B. J. *Fossil Horses* (Cambridge Univ. Press, 1992).
36. Favier, B. & Capecchi, M. R. Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of *hoxd-11*. *Development* **120**, 2187–2198 (1994).
37. Davis, A. P. & Capecchi, M. R. A mutational analysis of the 5' *HoxD* genes: Dissection of genetic interactions during limb development in the mouse. *Development* **122**, 1175–1185 (1996).
38. Favier, B. *et al.* Functional cooperation between the non-paralogous genes *Hoxa-10* and *Hoxd-11* in the developing forelimb and axial skeleton. *Development* **122**, 449–460 (1996).
39. Dollé, P. *et al.* Disruption of the *Hoxd-13* gene induces localized heterochrony leading to mice with neotenic limbs. *Cell* **75**, 431–441 (1993).
40. Favier, B., LeMeur, M., Chambon, P. & Dollé, P. Axial skeleton homeosis and forelimb malformations in *Hoxd-11* mutant mice. *Proc. Natl Acad. Sci. USA* **92**, 310–314 (1995).
41. Capecchi, M. R. Function of homeobox genes in skeletal development. *Ann. N. Y. Acad. Sci.* **97**, 34–37 (1996).
42. Wigglesworth, V. B. Evolution of insect wings and flight. *Nature* **246**, 127–203 (1973).
43. Budd, G. The morphology of *Opabinia regalis* and the reconstruction of the arthropod stem-group. *Lethaia* **29**, 1–14 (1996).
44. Hou, X. G. & Bergström, J. Cambrian lobopodians—ancestors of extant onychophorans? *Zool. J. Linn. Soc. Lond.* **114**, 3–19 (1995).
45. Simonetta, A. M. & Delle Cave, L. in *The Early Evolution of Metazoa and the Significance of Problematic Taxa* (eds Simonetta, A. M. & Conway Morris, S.) 189–244 (Cambridge Univ. Press, 1991).
46. Budd, G. A Cambrian gilled lobopod from Greenland. *Nature* **364**, 709–711 (1993).
47. Chen, J. Y., Ramsköld, L. & Zhou, G. Q. Evidence for monophyly and arthropod affinity of Cambrian giant predators. *Science* **264**, 1304–1308 (1994).
48. Carroll, S. B. Homeotic genes and the evolution of arthropods and chordates *Nature* **376**, 479–485 (1995).
49. Struhl, G. Genes controlling segmental specification in the *Drosophila* thorax. *Proc. Natl Acad. Sci. USA* **79**, 7380–7384 (1982).
50. Struhl, G. A homeotic mutation transforming leg to antenna in *Drosophila*. *Nature* **292**, 635–638 (1981).
51. Gibson, G. & Gehring, W. J. Head and thoracic transformations caused by ectopic expression of *Antennapedia* during *Drosophila* development. *Development* **102**, 657–675 (1988).
52. Stuart, J., Brown, S., Beeman, R. & Denell, R. A deficiency of the homeotic complex of the beetle *Tribolium*. *Nature* **350**, 72–74 (1991).
53. Averof, M. & Akam, M. *Hox* genes and the diversification of insect–crustacean body plans. *Nature* **376**, 420–423 (1995).
54. Vachon, G. *et al.* Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene. *Cell* **71**, 437–450 (1992).
55. Panganiban, G. *et al.* The development of crustacean limbs and the evolution of arthropods. *Science* **270**, 1363–1366 (1995).
56. Manton, S. M. *Mandibular Mechanisms and the Evolution of Arthropods* Vol. 247 (British Museum and Queen Mary College, London, 1964).
57. Wheeler, W. C., Cartwright, P. & Hayashi, C. Y. Arthropod phylogeny: a combined approach. *Cladistics* **9**, 1–39 (1993).
58. Boore, J. L., Collins, T. M., Stanton, D., Daehler, L. L. & Brown, W. M. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**, 163–165 (1995).
59. Cohen, S. M. & Jürgens, G. Proximal–distal pattern formation in *Drosophila*: cell autonomous requirement for *Distal-less* gene activity in limb development. *EMBO J.* **8**, 2045–2055 (1989).
60. Cohen, S. *et al.* *Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*. *Nature* **338**, 432–434 (1989).
61. Panganiban, G., Nagy, L. & Carroll, S. B. The development and evolution of insect limb types. *Curr. Biol.* **4**, 671–675 (1994).
62. Popadic, A., Rusch, D., Peterson, M., Rogers, B. T. & Kaufman, T. C. Origin of the arthropod mandible. *Nature* **380**, 395 (1996).
63. Jeram, A. J., Selden, P. A. & Edwards, D. Land animals in the Silurian: Arachnids and myriapods from Shropshire, England. *Science* **250**, 658–661 (1990).
64. Kukalová-Peck, J. *The Insects of Australia* 2nd edn (Cornell University Press, Ithaca, NY, 1991).
65. Snodgrass, R. *Principles of Insect Morphology* (McGraw-Hill, New York, 1935).
66. Kukalová-Peck, J. Origin and evolution of insect wings and their relation to metamorphosis, as documented from the fossil record. *J. Morphol.* **156**, 53–126 (1978).
67. Marden, J. H. & Kramer, M. G. Surface-skimming stoneflies: A possible intermediate stage in insect flight evolution. *Science* **266**, 427–430 (1994).
68. Cohen, B. *et al.* Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Development* **117**, 597–608 (1993).
69. Diaz-Benjumea, F. & Cohen, S. M. Interaction between dorsal and ventral cells in the imaginal disc directs wing development in *Drosophila*. *Cell* **75**, 741–752 (1993).
70. Carroll, S. B. *et al.* Pattern formation and eyespot determination in butterfly wings. *Science* **265**, 109–114 (1994).
71. Averof, M. & Cohen, S. M. Evolutionary origin of insect wings from ancestral gills. *Nature* **385**, 627–630 (1997).
72. Lee, J. J. *et al.* Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene *hedgehog*. *Cell* **71**, 33–50 (1992).
73. Tabata, T. *et al.* The *Drosophila* *hedgehog* gene is expressed specifically in posterior compartment cells and is a target of *engrailed* regulation. *Genes Dev.* **6**, 2635–2645 (1992).
74. Basler, D. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by *hedgehog* protein. *Nature* **368**, 208–214 (1994).
75. Posakony, L., Raftery, L. & Gelbart, W. Wing formation in *Drosophila melanogaster* requires *decapentaplegic* gene function along the anterior–posterior compartment boundary. *Mech. Dev.* **33**, 69–82 (1991).
76. Capdevila, J. & Guerrero, I. The *Drosophila* segment polarity gene *patched* interacts with *decapentaplegic* in wing development. *EMBO J.* **6**, 715–729 (1994).
77. Sanicola, M., Sekelsky, J., Elson, S. & Gelbart, W. M. Drawing a stripe in *Drosophila* imaginal discs: negative regulation of *decapentaplegic* and *patched* expression. *Genetics* **139**, 745–756 (1995).
78. Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range actions of a *Dpp* morphogen gradient. *Cell* **85**, 357–368 (1996).
79. Lecuit, T. *et al.* Two distinct mechanisms for long-range patterning by *Decapentaplegic* in the *Drosophila* wing. *Nature* **381**, 387–393 (1996).
80. Echelard, Y. *et al.* *Sonic hedgehog*, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417–1430 (1993).
81. Krauss, S., Concordet, J. P. & Ingham, P. W. A functionally conserved homolog of the *Drosophila* segment polarity gene *hh* is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* **75**, 1431–1444 (1993).
82. Chang, D. T. *et al.* Products, genetic linkage and limb patterning activity of a murine *hedgehog* gene. *Development* **120**, 3339–3353 (1994).
83. Riddle, R. D. *et al.* *Sonic hedgehog* mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1995).
84. Tickle, C. Genetics and limb development. *Dev. Genet.* **19**, 1–8 (1996).
85. Irvine, K. & Weischaus, E. *fringe*, a boundary-specific signaling molecule, mediates interactions between dorsal and ventral cells during *Drosophila* wing development. *Cell* **79**, 595–606 (1994).
86. Spreicher, S., Thomas, U., Hinz, U. & Knust, E. The *Serrate* locus of *Drosophila* and its role in morphogenesis or imaginal discs: control of cell proliferation. *Development* **120**, 535–544 (1994).
87. Kim, J., Irvine, K. & Carroll, S. Cell recognition, signal induction, and symmetrical gene activation at the dorsal–ventral boundary of the developing *Drosophila* wing. *Cell* **82**, 795–802 (1995).
88. Couso, J. P., Knust, E. & Martínez Arias, A. *Serrate* and *wingless* cooperate to induce *vestigial* gene expression and wing formation in *Drosophila*. *Curr. Biol.* **5**, 1437–1448 (1995).
89. Diaz-Benjumea, F. J. & Cohen, S. *Serrate* signals through *Notch* to establish a *Wingless*-dependent organizer at the dorsal/ventral compartment boundary of the *Drosophila* wing. *Development* **121**, 4215–4225 (1995).
90. Kim, J. *et al.* Integration of positional signals and regulation of wing formation and identity by *Drosophila* *vestigial* gene. *Nature* **382**, 133–138 (1996).
91. Todt, W. L. & Fallon, J. F. Development of the apical ectodermal ridge in the chick wing bud. *J. Embryol. Exp. Morphol.* **80**, 21–41 (1984).
92. Rodriguez-Estaban, C. *et al.* *Radical fringe* positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb. *Nature* **386**, 360–361 (1997).
93. Laufer, E. *et al.* Expression of *Radical fringe* in limb-bud ectoderm regulates apical ectodermal ridge formation. *Nature* **386**, 366–373 (1997).
94. Williams, J. A., Paddock, S. W. & Carroll, S. B. Pattern formation in a secondary field: A hierarchy of regulatory genes subdivides the developing *Drosophila* wing disc into discrete sub-regions. *Development* **117**, 571–584 (1993).
95. Couso, J. P., Bate, M. & Martínez-Arias, A. A *wingless*-dependent polar coordinate system in *Drosophila* imaginal discs. *Science* **259**, 484–489 (1993).
96. Parr, B. A. & McMahon, A. P. Dorsalizing signal *Wnt-7a* required for normal polarity of D-V and A-P axes of mouse limb. *Nature* **374**, 350–353 (1995).
97. Riddle, R. D. *et al.* Induction of the LIM homeobox gene *Lmx-1* by *Wnt-7a* establishes dorsoventral pattern in the vertebrate limb. *Cell* **83**, 631–640 (1995).
98. Vogel, A. *et al.* Dorsal cell fate specified by chick *Lmx1* during vertebrate limb development. *Nature* **378**, 716–720 (1995).
99. Raff, R. *The Shape of Life* (Univ. Chicago Press, 1996).
100. Panganiban, G. *et al.* The origin and evolution of animal appendages. *Proc. Natl Acad. Sci. USA* **94**, 5162–5166 (1997).
101. Wall, N. A. & Hogan, B. L. M. Expression of *bone morphogenetic protein-4* (*BMP-4*), *bone morphogenetic protein-7* (*BMP-7*), *fibroblast growth factor-8* (*FGF-8*) and *Sonic hedgehog* (*SHH*) during branchial arch development in the chick. *Mech. Dev.* **53**, 383–392 (1995).
102. Marigo, V., Scott, M. P., Johnson, R. L., Goodrich, L. V. & Tabin, C. J. Conservation in *hedgehog* signaling: induction of a chicken *patched* homolog by *Sonic hedgehog* in the developing limb. *Development* **122**, 1225–1233 (1996).
103. Roth, V. L. Homology and hierarchies: Problems solved and unresolved. *J. Evol. Biol.* **4**, 167–194 (1991).
104. Wagner, G. P. The origin of morphological characters and the biological basis of homology. *Evolution* **43**, 1157–1171 (1989).
105. Bolker, J. A. & Raff, R. A. Developmental genetics and traditional homology. *BioEssays* **18**, 489–494 (1996).
106. Carroll, R. L. *Vertebrate Paleontology* (Freeman, San Francisco, 1988).
107. Jarvik, E. *The Structure and Evolution of the Vertebrates* Vol. 1 (Academic, New York, 1980).
108. Jarvik, E. The Devonian tetrapod *Ichthyostega*. *Fossils and Strata* **40**, 1–213 (1996).
109. Fromental-Ramain, C. *et al.* Specific and redundant functions of the paralogous *Hoxa-9* and *Hoxd-9* genes in forelimb and axial skeleton patterning. *Development* **122**, 461–472 (1996).
110. Mortlock, D. P., Post, L. C. & Innis, J. W. The molecular basis of hypodactyly (Hd): a deletion in *Hoxa13* leads to arrest of digital arch formation. *Nature Genet.* **13**, 284–289 (1996).
111. Mortlock, D. P. & Innis, J. W. Mutation of *HOXA13* in hand–foot–genital syndrome. *Nature Genet.* **15**, 179–181 (1997).
112. Saunders, J. The proximo–distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–403 (1948).
113. Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
114. Lebedev, O. A. & Coates, M. I. The postcranial skeleton of the Devonian tetrapod *Tulerpeton curtum* Lebedev. *Zool. J. Linn. Soc.* **113**, 302–348 (1995).
115. Hou, X. G., Bergström, J. & Ahlberg, P. *Anomalacaris* and other large animals in the Lower Cambrian Chenjiang fauna of southwest China. *Geol. For. Forhandl.* **117**, 163–183 (1995).

**Acknowledgements.** We thank P. Ahlberg, G. Budd, A. C. Burke, M. Coates, A. Meyer, G. Panganiban, P. Sniegowski, D. Wake, R. S. Winters, L. Wolpert and members of our laboratories for their critiques of drafts of this manuscript. S.B.C. is an investigator of the HHMI. C.T. is supported by grants from the NIH and the American Cancer Society. N.S. is supported by grants from the NSF, from the National Geographic Society and from the Research Foundation of the University of Pennsylvania.

Correspondence and requests for materials should be addressed to N.S. (e-mail: nshubin@sas.upenn.edu).



# Deep winds on Jupiter as measured by the Galileo probe

David H. Atkinson\*, Andrew P. Ingersoll† & Alvin Seiff‡

\* Department of Electrical Engineering, University of Idaho, Moscow, Idaho 83844, USA

† Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

‡ San Jose State University Foundation, NASA Ames Research Center, Moffett Field, California 94035, USA

The Doppler Wind Experiment on the Galileo probe provided the first *in situ* data on wind speeds in Jupiter's atmosphere. Initial analysis<sup>1</sup> of the results indicated that wind speeds increase with depth, rather than decaying to zero below the cloud tops or remaining relatively constant as had previously been assumed<sup>2</sup>. But this earlier analysis was subject to several potential sources of error, as highlighted by the fact that wind speeds measured at the cloud tops did not seem to match those inferred from tracking clouds<sup>3</sup> in images obtained by the Voyager spacecraft. Here we report new analyses of the probe data that use a corrected treatment of the timing errors, adopt the measured<sup>4</sup> (rather than predicted) descent trajectory, and incorporate a new calibration of the instrumentation that takes into account the unexpectedly high temperatures encountered by the probe. We determine wind speeds at the cloud tops (700-mbar level) in the range 80–100 m s<sup>-1</sup>, in agreement with the results of cloud tracking; the speed increases dramatically between 1 and 4 bar, and then remains nearly constant at ~170 m s<sup>-1</sup> down to the 21-bar level. The increase in wind speed implies a latitudinal density gradient of 0.5% per degree in the 1–2 bar altitude range, but whether these winds are driven by internal heat or absorbed sunlight remains uncertain.

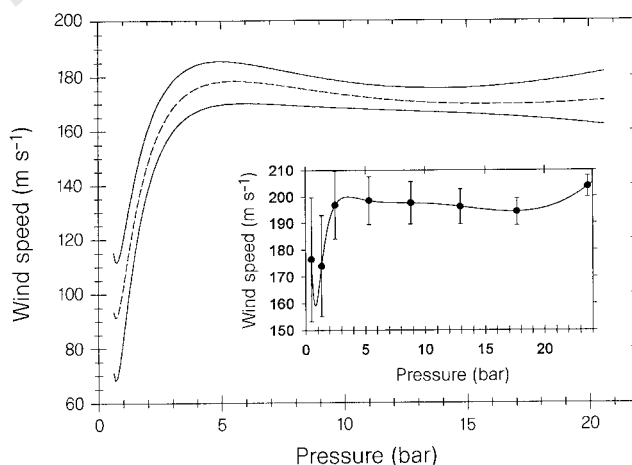
On arrival at Jupiter on 7 December 1995, the Galileo probe telemetered 57.5 minutes of atmospheric measurements to the orbiter. The probe telemetry frequency, measured on the orbiter, was altered from the transmitted frequency primarily by Doppler contributions from the orbiter trajectory, the probe trajectory (from descent and from the rotation of Jupiter) and the winds. When the nominal trajectories are removed from the measured frequency profile, the Doppler frequency residuals are attributed primarily to the effect of local zonal winds. To convert the Doppler residuals into a wind profile, an assumption must be made about the wind direction. At the latitude of probe entry, near 6.5° north, the meridional winds are an order of magnitude smaller than the zonal winds<sup>5</sup> and, by considerations of continuity, the vertical winds are expected to be three orders of magnitude smaller. It is therefore reasonable to project the frequency residuals onto the probe local east–west direction. Inversion of the frequency residuals, including accounting for the change in probe longitude (~0.45°) due to the integrated wind effect, yields a profile of zonal winds at the probe descent location. This profile, when combined with the nominal trajectory, provides a least-squares best fit to the measured frequency profile.

The early results from the Doppler Wind Experiment (DWE) presented several problems. First, the wind speed at the top of the clouds did not match the speed reported from tracking clouds in Voyager images<sup>3</sup> or from tracking the large-scale features at the time of probe entry<sup>6</sup>. The DWE gave a wind speed of  $175 \pm 25$  m s<sup>-1</sup> at cloud-top altitude. Tracking of the clouds at the latitude of the probe generally gives a value of 100–105 m s<sup>-1</sup>, although tracking small features sometimes gives speeds up to 160 m s<sup>-1</sup> (ref. 7). Second, by the end of the descent mission the probe interior reached a temperature of 150 °C, significantly higher than the maximum

expected operating temperature of 60 °C. The high temperatures might have caused unmodelled frequency shifts of the ultrastable oscillator leading to systematic errors in the derived velocities. Third, timing errors between the relay receiver hardware and the orbiter were modelled incorrectly. Last, the early DWE analysis used the predicted descent velocity of the probe rather than the measured descent velocity because the latter was not available owing to the higher-than-expected temperatures. As the descent velocity causes a major Doppler shift that must be removed before the Doppler shift due to the zonal winds can be measured, this was another potential source of error.

Based on the measured Doppler frequencies and improved analysis as described above, the retrieved zonal winds in the upper atmosphere are now in close agreement with cloud-top imaging from Voyager 2 (ref. 3). Near the 700-mbar level, the Doppler-measured winds are 90 m s<sup>-1</sup>. Between 1 bar and 4 bar the zonal wind speed increases rapidly, an effect also seen in the wind determinations by the accelerometers<sup>8</sup> of the Atmospheric Structure Instrument and the Earth-based Doppler measurements<sup>9</sup>. From 4 bar to 21 bar, the winds remain relatively constant near 170 m s<sup>-1</sup> (Fig. 1).

An error envelope spanning a range of possible wind profiles is shown in Fig. 1; it includes descent velocities that vary from 95% to 105% of nominal at each time point, probe entry and initial descent longitudes within  $\pm 0.21^\circ$  ( $3\sigma$ ) of nominal, and assuming the



**Figure 1** The Doppler wind measurements (main figure) now incorporate a number of significant changes from the preliminary results of Atkinson *et al.*<sup>1</sup> (inset). The new wind speeds were calculated using the probe descent velocity profile obtained from measurements of pressure and temperature by the Atmospheric Structure Instrument, the assumption of hydrostatic equilibrium, and the equation of state. Descent velocities are now modelled to better than 5% (~1–5 m s<sup>-1</sup>) throughout the descent, and the probe entry and initial descent longitude is known to  $\pm 0.21^\circ$  ( $3\sigma$ ). Periodic null frequency measurements are thought to be caused by imperfections in the relative timing of the relay receiver hardware and orbiter clocks. When unaccounted for, the timing mismatch between the relay receiver hardware and the orbiter ultrastable oscillator (USO) causes periodic discontinuities of ~7 Hz in the measured frequencies. Our treatment of the timing errors involves averaging through, as opposed to complete removal of, the discontinuities, resulting in a net reduction of the frequency residuals by ~50 Hz and a subsequent decrease in the retrieved deep zonal winds of 30 m s<sup>-1</sup>. The probe USO, qualification-tested to 60 °C and acceptance-tested to 50 °C, in fact experienced temperatures nearing 150 °C at the end of probe transmission. Thermal tests at NASA Ames Research Center on several flight spare USOs have recently concluded; the high-temperature behaviour of the USOs has been characterized and incorporated into the probe oscillator frequency model. The error envelope (solid lines) of wind speeds is based on worst-case descent velocity variations ( $\pm 5\%$ ), probe longitude error ( $\pm 0.21^\circ$ ,  $3\sigma$ ) and USO thermal response corrections (none/maximum).

minimum (no thermal corrections) and maximum thermal response models of the probe ultrastable oscillator. Given our current understanding of the high-temperature response of this oscillator, it is now seen that the tendency reported previously<sup>1</sup>—towards increasing winds in deepest regions explored by the probe—has disappeared.

The above-mentioned wind measurements by the Atmospheric Structure Instrument's accelerometers<sup>8</sup> are particularly reassuring in that they are obtained by an entirely different technique that does not depend on radio tracking, the stability of a crystal oscillator, or the subtraction of other contributing velocities. They are of low velocity resolution and limited accuracy, however, and therefore do not give precise wind speeds. The radio tracking measurements from Earth (by the Very Large Array) have a much better viewing angle than that available from the overhead Galileo orbiter, but were significantly affected by the unknown frequency offset of the probe ultrastable oscillator and the enormous transmission distance leaving only a faint signal at Earth which must be extracted from the noise<sup>9</sup>. It is remarkable that these three techniques agree on the essential description of the winds, leaving little doubt that they have been characterized.

Although there was speculation about whether the winds were shallow or deep, very few had predicted that the winds would increase with depth. By 'shallow' one usually means 'above the base of the water cloud', that is, above the 5-bar level. This is the region where sunlight is absorbed and latent heat is released, so it was assumed that if the winds drew their energy from these sources they should decay with depth below the cloud base<sup>2,10–12</sup>. On the other hand, if the winds drew their energy from internal heat, they might continue down indefinitely into the fluid interior of Jupiter<sup>13–15</sup>. Only one analysis<sup>16</sup> predicted that the winds would increase with depth. The DWE has shown that the winds are 'deep'. But it is not clear that this means that they are powered by internal heat; neither is it clear why the winds increase with depth at the 1–4 bar level.

In the atmospheres of rotating planets, the variation of zonal wind with depth is proportional to the variation of density with latitude. The density gradient is taken at constant pressure. If the gradient is zero, the fluid is said to be barotropic. If the interior of a fluid planet is barotropic, then each cylinder, concentric with the axis of rotation, will rotate as a solid body. These statements, proved nearly a century ago<sup>17</sup>, are valid when forcing and dissipation are small. The forcing that maintains the winds may be confined to the surface layers or it may be distributed throughout the fluid. Thus we cannot say with certainty that the zonal winds are powered by internal heat.

The fluid is clearly not barotropic in the 1–4 bar pressure range where zonal wind is increasing with depth. The increase implies a density decrease of 0.5% per degree of latitude at the 1–2 bar levels. This is three to four times larger than, but of the same sign as, that inferred from the temperature gradients observed at the 150- and 270-mbar levels<sup>18</sup>. The latter have been interpreted as a sign of upwelling to the south, and downwelling to the north, of the probe site. The DWE results suggest that the upwellings and downwellings extend at least to the 4-bar level.

The constancy of the wind in the 4–21 bar range implies that the fluid there is barotropic. This is expected, as sunlight does not penetrate<sup>19</sup> and latent heat is not released at these levels. Convection and radiation, the other modes of heat transfer, are likely to maintain a barotropic state<sup>20</sup>. Thus the 170 m s<sup>-1</sup> winds measured by the DWE probably extend well into the fluid interior of Jupiter. This result is qualitatively consistent with studies of cloud-top winds, whose vorticity can be used to infer the winds at depth<sup>21</sup>. Magnetic fields<sup>22</sup>, phase transitions, and small gradients of temperature can alter the barotropic state, so the zonal winds do not necessarily extend throughout the planet. But the DWE result shows that the winds are deep, and this has important implications for atmospheric dynamics. □

Received 7 March; accepted 3 June 1997.

1. Atkinson, D. H., Pollack, J. B. & Seiff, A. Galileo Doppler measurements of the deep zonal winds at Jupiter. *Science* **272**, 842–843 (1996).
2. Pollack, J. B., Atkinson, D. H., Seiff, A. & Anderson, J. D. Retrieval of a wind profile from the Galileo probe telemetry signal. *Space Sci. Rev.* **60**, 143–178 (1992).
3. Limaye, S. S. Jupiter: new estimates of the mean zonal flow at the cloud level. *Icarus* **65**, 335–352 (1986).
4. Seiff, A. *et al.* Structure of the atmosphere of Jupiter: Galileo probe measurements. *Science* **272**, 844–845 (1996).
5. Ingersoll, A. P. *et al.* Interaction of eddies and mean zonal flow on Jupiter as inferred from Voyager 1 and 2 images. *J. Geophys. Res.* **86**, 8733–8743 (1981).
6. Orton, G. *et al.* Earth-based observations of the Galileo probe entry site. *Science* **272**, 839–840 (1996).
7. Beebe, R. E., Simon, A. A. & Huber, L. F. Comparison of Galileo probe and Earth-based translation rates of Jupiter's equatorial clouds. *Science* **272**, 841 (1996).
8. Seiff, A. *et al.* Wind speeds measured in the deep jovian atmosphere by the Galileo probe accelerometers. *Nature* **388**, 650–652 (1997).
9. Folkner, W. M. *et al.* Earth-based radio tracking of the Galileo probe for Jupiter wind estimation. *Science* **275**, 644–646 (1997).
10. Hess, S. L. & Panofsky, H. A. in *Compendium of Meteorology* (ed. Malone, T. F.) 391–398 (Am. Meteorol. Soc., Boston, 1951).
11. Ingersoll, A. P. & Cuzzi, J. N. Dynamics of Jupiter's cloud bands. *J. Atmos. Sci.* **26**, 981–985 (1969).
12. Williams, G. P. Jovian dynamics. Part I: vortex stability, structure, and genesis. *J. Atmos. Sci.* **53**, 2685–2734 (1996).
13. Ingersoll, A. P. & Cuong, P.-G. Numerical model of long-lived jovian vortices. *J. Atmos. Sci.* **38**, 2067–2076 (1981).
14. Busse, F. H. A simple model of convection in the jovian atmosphere. *Icarus* **29**, 255–260 (1976).
15. Ingersoll, A. P. & Pollard, D. Motion in the interiors and atmospheres of Jupiter and Saturn: scale analysis, anelastic equations, barotropic stability criterion. *J. Atmos. Sci.* **52**, 62–80 (1982).
16. Dowling, T. E. Estimate of Jupiter's deep zonal-wind profile from Shoemaker–Levy 9 data and Arnold's second stability criterion. *Icarus* **117**, 439–442 (1995).
17. Poincaré, H. Sur la Précession des corps déformables. *Bull. Astron.* **27**, 231–356 (1910).
18. Gierasch, P. J., Conrath, B. J. & Magalhaes, J. O. Zonal mean properties of Jupiter's upper troposphere from Voyager infrared observations. *Icarus* **67**, 456–483 (1986).
19. Sromovsky, L. A. *et al.* Solar and thermal radiation in Jupiter's atmosphere: initial results of the Galileo probe net flux radiometer. *Science* **272**, 851–854 (1996).
20. Ingersoll, A. P. & Porco, C. C. Solar heating and internal heat flow on Jupiter. *Icarus* **35**, 27–43 (1978).
21. Dowling, T. E. & Ingersoll, A. P. Jupiter's great red spot as a shallow water system. *J. Atmos. Sci.* **46**, 3256–3278 (1989).
22. Kirk, R. L. & Stevenson, D. J. Hydromagnetic constraints on deep zonal flow in the giant planets. *Astrophys. J.* **316**, 836–846 (1987).

**Acknowledgements.** We thank the Galileo Project and NASA's Planetary Atmospheres Program. This work was supported by NASA.

Correspondence and requests for materials should be addressed to D.H.A. (e-mail: atkinson@maxwell.ee.uidaho.edu).

## Wind speeds measured in the deep jovian atmosphere by the Galileo probe accelerometers

**A. Seiff<sup>†</sup>, R. C. Blanchard<sup>‡</sup>, T. C. D. Knight<sup>§</sup>, G. Schubert<sup>||</sup>, D. B. Kirk<sup>§</sup>, D. Atkinson<sup>†</sup>, J. D. Mihalov<sup>\*</sup> & R. E. Young<sup>\*</sup>**

<sup>\*</sup> NASA-Ames Research Center, Moffett Field, California 94035, USA

<sup>†</sup> Department of Meteorology, San Jose State University, San Jose, California, USA

<sup>‡</sup> NASA Langley Field, MS-107, Hampton, Virginia 23665, USA

<sup>||</sup> Department of Earth & Space Science, UCLA, Los Angeles, California 90024, USA

<sup>§</sup> Department of Electrical Engineering, University of Idaho, Moscow, Idaho 83844, USA

The atmosphere of Jupiter has a complex circulation which, until recently, has been observable only at the cloud tops<sup>1,2</sup>; the mechanisms driving the winds, and the nature of the interior circulation, remained unknown<sup>3</sup>. Recent analyses<sup>4–6</sup> of the radio signal from the Galileo probe, obtained during its descent into the jovian atmosphere, have suggested a vigorous interior circulation below the 4-bar level. Here we report an independent measurement of the winds below the cloud tops, making use of the data obtained by the two accelerometers on the descending probe. We find evidence for two distinct wind regimes, in general agreement with the Doppler radio measurements: a region of wind shear between 1 and 4 bar, where the wind speed increases dramatically with depth; and then a region of constant high-velocity winds down to at least the 17-bar level.

We present here acceleration data and their interpretation to

§ Retired.

define zonal wind velocities. Two identical sensors,  $z_1$  and  $z_2$ , aligned with the probe axis of symmetry (the  $z$  axis) returned data shown in Fig. 1 during 59 minutes of probe descent on the parachute. The data are plotted as functions of atmospheric pressures measured by the Atmosphere Structure Instrument<sup>7,8</sup>. In a frame of reference moving with the wind, the  $z$  axis and the descent velocity were close to vertical during parachute descent. As expected, the measured accelerations agree approximately with Jupiter's acceleration due to gravity,  $g$ . Measured accelerations were initially greater than  $g$ , then fell below  $g$  at atmospheric pressure  $p > 2$  bar. For  $p > 4$  bar, the data are roughly parallel to the  $g$  curve. Data from the two sensors are generally quite similar.

Deviations of the measured accelerations,  $a_z$ , from the acceleration due to gravity are explained by inputs other than gravity, of which there were two, as represented in the following equation:

$$a_z = a_g - \omega^2 R \cos^2 \theta + dw/dt + L(\pi\delta/T)^2 \quad (1)$$

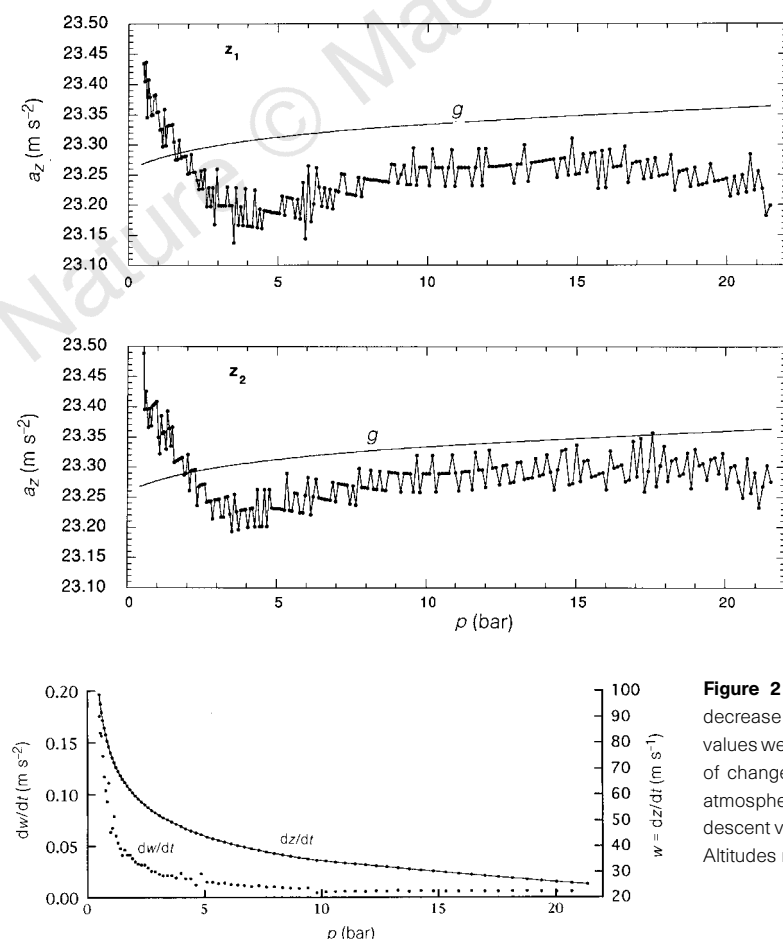
Here, following the convention used in geodesy,  $a_g$  is the gravitational acceleration (uncorrected for centrifugal acceleration),  $\omega$  is the planetary rotational velocity ( $\text{rad s}^{-1}$ ),  $R$  is the probe radial distance from the centre of the planet, and  $\theta$  is the jovian latitude. Thus  $\omega^2 R \cos^2 \theta$  is the radial component of centrifugal acceleration due to planetary rotation. Together, the first two terms comprise  $g$ , the gravity acceleration. Variation of the gravitational acceleration  $a_g$  with radius was represented simply by  $a_g = a_{g0}(R_0/R)^2$  where subscript 0 identifies the 1-bar level. The two additional terms are: first,  $dw/dt$ , the time rate of change of descent velocity, which is the real deceleration of the probe resulting from increasing atmospheric density; and second,  $L(\pi\delta/T)^2$ , the mean acceleration over one oscillation cycle of the probe swinging as a pendulum beneath the parachute (here  $L$  is the pendulum length,  $\delta$  is the swinging amplitude and  $T$  is the period). Both terms

add to tension in the parachute cord, and increase the acceleration felt by the probe. Analogously, the restraining force needed to hold the accelerometer test mass in fixed position, which is the measure of the acceleration, is increased. The observed swinging motion executed  $\sim 3$  cycles in 16 s with  $L = 13.9$  m,  $\delta = 2.2^\circ$  and  $T = 4.9$  s (ref. 8): this defines a mean pendulum acceleration of  $0.0084 \text{ m s}^{-2}$ . The real deceleration of the probe,  $dw/dt$ , was derived by differentiating the descent velocities, shown in Fig. 2. It is evident that this term explains the excess of measured acceleration over  $g$  early in the descent (Fig. 1).

From equation (1), we expect  $a_z - dw/dt - L(\pi\delta/T)^2$  to be a measure of  $g$ . However, the data so adjusted do not agree with  $g$ , unless allowance is made for additional centrifugal acceleration associated with winds (Fig. 3). With zonal winds present,  $\omega = \omega_j + u_w/R \cos \theta$  where  $\omega_j$  is Jupiter's rotational velocity and  $u_w$  is wind velocity, and two new terms appear on the right-hand side of equation (1),  $2\omega_j u_w \cos \theta + u_w^2/R$ . Term magnitudes at  $u_w = 100 \text{ m s}^{-1}$  are  $2\omega_j u_w \cos \theta = 0.03495 \text{ m s}^{-2}$  and  $u_w^2/R = 0.00014 \text{ m s}^{-2}$ . Curves of  $g$  corrected for these terms are shown in Fig. 3 for wind velocities of 100, 200 and  $300 \text{ m s}^{-1}$ .

In relation to these curves, the data indicate the presence and the magnitude of the winds. They clearly indicate a shear layer between  $\sim 1.5$  and 4 bar terminating in high velocities ( $\sim 200 \text{ m s}^{-1}$ ) which then persist below the 4-bar level for the remainder of the descent. The two main characteristics of the Doppler winds, the existence of a shear layer from 1.5 to 4 bar and essentially constant deep atmosphere winds, are thus confirmed. This is an important result as it gives information about the planet's internal circulation.

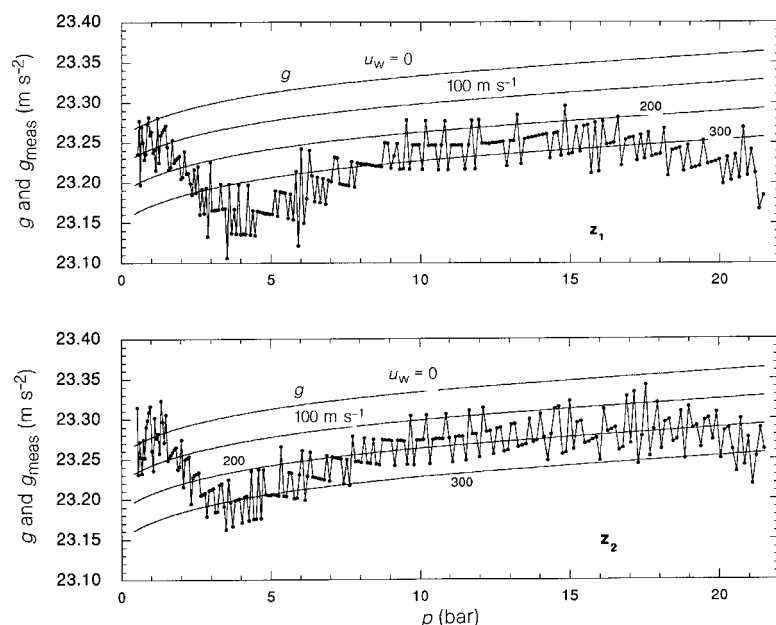
The wind magnitude definition by the accelerometers is limited by the relatively coarse digital resolution of the measurement, with 1 count equivalent to  $87.4 \text{ m s}^{-1}$  of zonal velocity. A 1-count uncertainty is rather typical of digital data. Data taken during cruise



**Figure 1** Axial accelerations  $a_z$  of the Galileo probe (data points) measured during parachute descent ( $p$  is atmospheric pressure). Data are shown from accelerometers  $z_1$  (top panel) and  $z_2$  (bottom panel). Each value plotted is the average of 512 readings in a 16-s sampling interval. The scatter is  $\pm 1$  count about the local mean, with 1 count =  $0.031 \text{ m s}^{-2}$ . The gravity curve  $g$ , which is corrected for planetary spin, varies perceptibly with radial distance to the planet's centre, and hence with pressure. For the measured gravitational field<sup>10</sup> and  $\omega_j = 0.00017585 \text{ rad s}^{-1}$ , the gravitational and centrifugal accelerations at the 1-bar level at the entry site are  $25.456$  and  $2.178 \text{ m s}^{-2}$ , respectively, giving  $a_g = a_{g0} = 23.278 \text{ m s}^{-2}$ . At the probe descent latitude,  $\theta = 6.42^\circ$ ,  $R_0 = 71,334 \text{ km}$  at the 1-bar level, corresponding to an equatorial radius of  $71,400 \text{ km}$  (ref. 10).

**Figure 2** The descent velocity ( $dz/dt$ ) and the probe deceleration ( $dw/dt$ ) decrease with increasing atmospheric pressure ( $p$ ) in parachute descent. These values were derived, with the assumption of hydrostatic equilibrium, from the rate of change of pressure with time indicated by pressure data from the Galileo atmosphere structure experiment. These data also define the contribution of descent velocity to line-of-sight velocity required in the Doppler wind experiment. Altitudes relative to the 1-bar level were from  $+16$  to  $-129 \text{ km}$ .





**Figure 3** Data from accelerometers  $z_1$  (top panel) and  $z_2$  (bottom panel), reduced by the real deceleration of the probe ( $dw/dt$ ) and probe swinging accelerations, compared with gravity accelerations modified for several zonal wind speeds  $u_w$ . The comparisons indicate wind speeds increasing between the 0.5- and the 4-bar levels, then becoming constant at  $\sim 200 \text{ m s}^{-1}$ . In each panel, the uppermost  $g$  curve is for zero wind. The three lower curves indicate accelerations of gravity experienced by a probe moving with the winds at 100, 200 and  $300 \text{ m s}^{-1}$ .

checks and on the ground indicate that system offsets (sensors plus electronics) were stable within  $\sim 1$  count. The sensor offsets were shown—by analogue calibrations in 1983–85—to be stable within 0.15 counts ( $z_1$ ) and 0.05 counts ( $z_2$ ) over a period of 18 months. Self checks of the electronics during descent and the data themselves indicate that the offset was stable within  $\sim 1$  count during the descent. There were, however, two disturbing influences: just before descent, the probe experienced the 228g deceleration of entry. Immediately thereafter, they were exposed to a probe internal environmental temperature swing from 0 to  $-35^\circ\text{C}$  and finally to  $+115^\circ\text{C}$ . The sensors were designed and calibrated for the operating range from  $-20$  to  $+50^\circ\text{C}$ . Outside this range, it was necessary to extrapolate the calibrations. Below 2 bar and from 9 to 17 bar data were taken at sensor temperatures within the calibration range.

To correct for the large and rapid variation in probe internal temperature, offset or zero reading shifts were calculated, based on data taken in 1983 and 1984 during calibration temperature sweeps at rates of  $\sim 1^\circ\text{C min}^{-1}$ . The  $z_2$  correction was 1.6 counts after passing through the temperature minimum while the  $z_1$  correction was a maximum of 0.3 counts. The corrected data indicate wind speeds of  $230 \pm 30 \text{ m s}^{-1}$  ( $z_1$ ) from 9 to 16 bar, and  $\sim 180 \text{ m s}^{-1}$  ( $z_2$ ) between 8 and 18 bar. The  $z_2$  data parallel the  $g$  contour, implying constant wind speed. The Doppler wind magnitude in the deep atmosphere,  $175 \text{ m s}^{-1}$  with formal uncertainty  $\sim \pm 5 \text{ m s}^{-1}$ , also agrees with the  $z_2$  accelerometer determination. The disagreement of the two sensors is within the 1-count uncertainty.

From 0.6 to 1.5 bar, the present data indicate relatively constant winds at speeds somewhat smaller than the cloud-top winds measured from Voyager images ( $90\text{--}150 \text{ m s}^{-1}$ )<sup>1,2</sup>. The  $z_1$  data indicate velocities generally  $< 100 \text{ m s}^{-1}$  whereas the  $z_2$  data indicate near zero velocity on average. (We feel that the sensor zero reading is in question there.) These data are, however, within 1 count of agreeing with previously observed cloud-top winds.

The general agreement between the determinations of the winds by the accelerometers and by Doppler tracking is highly significant. Both measurements indicate two distinct wind regimes: the region of wind shear with winds increasing from 1 to 4 bar; and the region of constant, high-velocity winds ( $\sim 200 \text{ m s}^{-1}$ ) from 4 to at least 17 bar. Below the 4-bar level, the measured winds are probably driven by deep-seated processes associated with the outflow of internal heat, whereas above the 4-bar level, 'surface' effects modify the deep winds. To modify the winds vertically, meridional temperature gradients are needed.<sup>5</sup> These can be provided by the meridional

differences in cloud cover, of which the belts and zones are the most prominent<sup>9</sup>: these cloud cover variations are accompanied by differences in solar heating and heat of condensation, and so are a source of meridional temperature contrast. □

Received 10 March; accepted 25 June 1997.

1. Limaye, S. *et al.* Jovian winds from Voyager 2, Part I: zonal mean circulation. *J. Atmos. Sci.* **39**, 1413–1432 (1982).
2. Beebe, R. F., Simon, A. A. & Huber, L. F. Comparison of Galileo Probe and Earth-based translation rates of Jupiter's equatorial clouds. *Science* **272**, 841 (1996).
3. Pollack, J. B., Atkinson, D. H., Seiff, A. & Anderson, J. D. Retrieval of a wind profile from the Galileo Probe telemetry signal. *Space Sci. Rev.* **60**, 143–178 (1992).
4. Atkinson, D. P., Pollack, J. B. & Seiff, A. Galileo Doppler measurements of the deep zonal winds at Jupiter. *Science* **272**, 842–843 (1996).
5. Atkinson, D. H., Ingersoll, A. P. & Seiff, A. Deep winds on Jupiter as measured by the Galileo probe. *Nature* **388**, 649–650 (1997).
6. Folkner, W. M. Earth-based radio tracking of the Galileo Probe for Jupiter wind estimation. *Science* **275**, 844–845 (1997).
7. Seiff, A. & Knight, T. C. D. K. The Galileo Probe atmosphere structure instrument. *Space Sci. Rev.* **60**, 203–232 (1992).
8. Seiff, A. *et al.* Structure of the atmosphere of Jupiter: Galileo Probe measurements. *Science* **272**, 844–845 (1996).
9. Ingersoll, A. P. & Cuzzi, J. N. Dynamics of Jupiter's cloud bands. *J. Atmos. Sci.* **26**, 981–985 (1969).
10. Anderson, J. D. in *Jupiter* (ed. Gehrels, J.) 113–121 (Univ. Arizona Press, Tucson, 1976).

Correspondence should be addressed to A.S. at MS 245-1, NASA-Ames Research Center, Moffett Field, California 94035, USA (e-mail: aseiff@mail.arc.nasa.gov).

## Spontaneous polarization in dense hydrogen

**B. Edwards & N. W. Ashcroft**

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853-2501, USA

More than six decades have passed since Wigner and Huntington<sup>1</sup> proposed that hydrogen might form a solid metallic phase at high density with characteristics similar to the alkali metals. This possibility has been investigated using the diamond-anvil cell to compress the crystalline state of molecular hydrogen<sup>2</sup>, but there is still no definitive evidence for a dense, low-temperature metallic state. Below 140 K, solid hydrogen undergoes a transition at about 1.5 million atmospheres between two orientationally ordered states. The intermolecular vibrational mode (the vibron) shifts to a lower frequency at this transition<sup>3,4</sup>, and becomes strongly infrared-active<sup>5</sup>. So far as is known, hydrogen remains in this phase to the highest pressures yet reached. Here we report

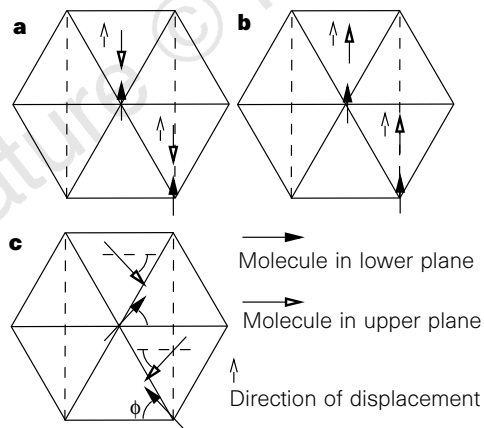
first-principles calculations of the structure of this phase using electronic density-functional theory. We find that it develops a spontaneous polarization at around ninefold compression relative to the volume at 1 atmosphere and that there is a corresponding movement of proton pairs away from their ideal lattice sites. Such behaviour can explain why the vibron becomes infrared-active, and rationalizes the direction and mass-dependence (in experiments on deuterium) of the shift of the vibron frequency. In the polarized state, the previously decreasing bandgap widens again, and so its appearance might delay the transition to the elusive metallic state.

The state we find can be viewed as a charge density wave reflecting a breaking of electronic symmetry. Arguments for its onset can be given at two levels. First we look at dielectric instability. We consider an assembly of  $N$  molecules of hydrogen represented as static objects initially with isotropic linear polarizability  $\alpha$ , each fixed at the sites of a lattice of macroscopic volume  $V$ . We fix attention on one of these molecules and, beginning with an infinitesimal external field, let all others acquire small identical dipoles  $\mathbf{d}$ , producing a field  $\mathbf{E} = (N/V)f\mathbf{d}$  at each site (where  $f$  is dependent on the details of the structure). The factor  $f$  can vary considerably, acquiring even negative values for some configurations, but for energetically favourable arrangements of the molecules  $f \approx 5$  (ref. 6). The action of this field is to polarize the molecule at the selected site, and the dipole moment  $\mathbf{d}'$  induced there is then  $\alpha\mathbf{E}$  or equivalently  $\mathbf{d}' = (3\alpha f/8\pi r_s^3 a_0^3)\mathbf{d}$ , where  $a_0$  is the Bohr radius and  $r_s$  is the standard linear measure of average electronic density, given by  $4\pi r_s^3 a_0^3/3 = V/(2N)$  (the factor of two merely accounts for two electrons per molecule). It follows that, if  $r_s < [3\alpha f/(8\pi a_0^3)]^{1/3}$  and the polarizability is held constant, then a route is identified by which the dipole moments of all molecules and the local field at them can appear to collectively rise, ostensibly without bound. Phrased in the language of molecular field theory, this argument is entirely of a mean-field character, and it simply shows that for  $r_s < [3\alpha f/(8\pi a_0^3)]^{1/3}$  the ground state may be an ordered state, in

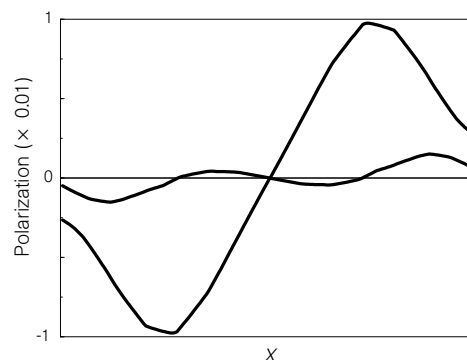
this case one of spontaneous polarization. These arguments, originating with Herzfeld<sup>7</sup> and Goldhammer<sup>8</sup>, identify from the insulating side the possible onset of an electronic transition with increase in density. The principal assumptions, that the polarizability remains both constant and isotropic and that higher multipoles can be ignored, are not without consequence. Nevertheless, if the average polarizability of the free hydrogen molecule ( $5.6a_0^3$ ) is taken as a guide, a critical  $r_s$  value of 1.49 is predicted (experimentally, the onset of pronounced infrared activity in dense hydrogen is at  $r_s \approx 1.5$ , or ninefold compression; ref. 5). The polarizability is, however, far from isotropic and as an alternative manifestation of electronic distortion we may consider a self-sustaining polarization possibly accompanied by reorientation and displacement of the molecules.

Though some aspects of the description just given are intuitively reasonable, the argument clearly omits (for example through the assumption of constant polarizabilities) the significant effects associated with the appreciable overlap of molecular electronic wavefunctions that necessarily accompanies compression. By application of pressure hydrogen, which can be thought of initially as a divalent entity, can be brought to conditions where band-widths surpass 20 eV. Any notion of spontaneous polarization must therefore be examined first from the perspective of one-electron theory and then its subsequent correction when residual Coulomb effects are restored to the one-electron viewpoint. Proton zero-point motion is of considerable importance to this picture and will be taken up below.

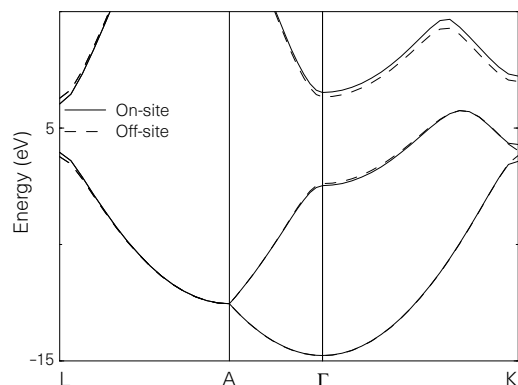
Guided by the range of potentially interesting densities suggested above, we turn to our second reason for proposing an electronic instability: it is revealed by the results of our very detailed electronic structure and total-energy calculations on dense hydrogen within the local-density approximation of density-functional theory. The algorithm we use is that of Teter *et al.*<sup>9</sup>, and a pseudopotential with a matching radius of  $0.5a_0$  has been constructed to take care of the singular electron–proton interaction. A plane-wave cut-off of 80 Ry



**Figure 1** Hexagonal close-packed structures of dense hydrogen considered in this Letter (unit vectors in the hexagonal plane have length  $a$ , and perpendicular to it  $c$ ). The bond lengths,  $\sim 1.45a_0$  in all three cases, do not change significantly as the molecules move off-site. The  $c:a$  ratio is 1.58. The heads of the large arrows indicate the uppermost ends of the molecules, and the filled and unfilled arrows represent molecules in planes separated by  $c/2$ . **A**,  $Cmc2_1$  (ref. 10). Hydrogen molecules are tilted out of the hexagon plane (shown) by the angle  $\theta = 30^\circ$ . The displacements correspond to a horizontal sliding of the upper layer along the direction indicated by the small arrows. **B**,  $C2/m$  (ref. 10). Here  $\theta = 20^\circ$ , and the molecules in both layers have identical orientations. The dipole moments of molecules in alternate layers are directed oppositely so the arrangement is antiferroelectric. **C**,  $Pca2_1$  (ref. 11). There are four molecules per primitive unit cell, with  $(\theta, \phi) = (27^\circ, 47^\circ)$ . There is no displacement away from exact lattice sites in  $Pca2_1$ .



**Figure 2** Charge asymmetry along a molecular bond associated with spontaneous polarization, for the structure  $Cmc2_1$  at  $r_s = 1.47$ . Polarization is defined as  $(\rho(x) - \rho(-x))/2\rho_0$ , with  $\rho(x)$  the electron density,  $x = 0$  defined as the midpoint of the molecule, and  $\rho_0$  the maximum electron density in the ground state of the hydrogen atom. The large positive peak on the right therefore indicates a fractional excess of electron charge over that of the left side, which has a corresponding deficit. The smaller-amplitude curve shows the polarization before the molecules have moved off-site; this is associated with the overall symmetry imposed by the crystal structure, which lacks inversion symmetry. We note that this on-site polarization is an order of magnitude too small to account for the infrared activity observed in phase III. It is, however, of the right order to explain the smaller infrared activity found in phase II (ref. 12), which we thus interpret as hindered but polarized only weakly (by the crystal field).



**Figure 3** Energy bands before (full curves) and after (dashed curves) the onset of the state of spontaneous polarization in  $Cmc2_1$  at  $r_s = 1.47$ . The overall effect is to further widen (by  $\sim 0.2$  eV) a previously declining indirect bandgap, thereby delaying the expected onset of the metal-insulator transition. We note that the valence bandwidth is  $\sim 20$  eV, and that the conduction band minimum is at L.

and an  $8 \times 8 \times 8$  grid of wave-vectors  $\mathbf{k}$  in the Brillouin zone of the four-molecule orthorhombic cell (Fig. 1) were used throughout; all of these parameters were carefully checked for their effects on convergence, and at the end of each such calculation we examine the disposition of electronic charge in the chosen cells. Crucial to our method is the fact that the positions of the protons were relaxed to find a local energy minimum using forces resulting from fully converged *ab initio* calculations at each step of proton displacement. This has not been common practice in published *ab initio* calculations of the properties of dense solid hydrogen, the usual route being to constrain the centres of mass of the proton pairs to lie exactly on the lattice sites of the structures under consideration. As we have repeatedly found, removal of this constraint leads to very different results.

These procedures have been applied to several structures suggested as energetically competitive by earlier work (Fig. 1)<sup>10,11</sup>. For each, we start with the protons at or near the exact sites of the chosen structure but then fully relax their coordinates to find a local energy minimum. Our principal result is that in certain structures and for  $r_s \leq 1.5$  the molecules move away from the lattice sites by as much as  $0.2a_0$ , and a significantly asymmetric distribution of electronic charge develops; the system is spontaneously polarized. This asymmetry of the electronic charge on each molecule is an order of magnitude larger than the weak permanent asymmetries generally expected in those structures in which the centres of the molecules are not at sites of inversion symmetry (Fig. 2). The bond lengths,  $\sim 1.45a_0$  in all three structures, do not change significantly as the molecules move off-site. Displacement from exact lattice sites and the dramatic increase in polarization are found for several, but not all, structures (Fig. 1). In particular, among the hexagonal-based structures considered here, it does not occur in the  $Pca2_1$  structure<sup>11</sup>, which is the lowest-energy structure so far found in which molecular centres lie on hexagonal close-packed sites. When this constraint is released the energy of the  $Cmc2_1$  structure (Fig. 1a) decreases to below that of  $Pca2_1$ . The energy of the  $C2/m$  structure also decreases when the constraint is removed, although it remains slightly above the energy of the other two. In this structure there are two sublattices with oppositely directed polarizations in a classic antiferroelectric arrangement. The energy differences among these structures are small and therefore it cannot be stated with certainty that any reported structure will necessarily represent a global energy minimum; nevertheless the transition discussed above does lower the energy of those structures that undergo it. So, permitting the molecules the freedom to move away from exact lattice sites seems to be a necessity here.

Below  $r_s \leq 1.5$  the growing charge asymmetry is initially at the per cent level, and is in good accord with the measured infrared absorptivities<sup>5</sup>, which we attribute to this spontaneous polarization. The arbitrariness of phase in the charge density wave description of this effect is reflected in a periodic system in a lack of uniqueness in the definition of the polarization. Here, for a molecular system, the dipoles are located, as expected, at molecular centres. The subsequent inclusion of proton zero-point motion modifies this asymmetry. First, variation in molecular polarizability with any changes in bond length can lead to a charge asymmetry larger than that found here for fixed configurations of the protons<sup>12</sup>; this will therefore lead to an increase of the effective polarization, which is related to the intensity of the infrared absorption. This effect is estimated to increase the polarization by as much as a factor of  $\sim 2$  (ref. 12), so that our estimated effective polarization is  $\sim 3\%$  (close to the experimental value<sup>12</sup>). Second, inclusion of the librational modes acts to decrease the average charge asymmetry, as discussed below. Other modes will also affect the charge asymmetry somewhat in a system that possesses such large-amplitude zero-point motion. The polarization also affects the band structure, the result being a widening of the indirect energy gap (Fig. 3) which, before the onset of the transition, was steadily declining with increase in pressure.

These effects are exactly what is expected from the formation of a charge density wave (the Peierls transition being the standard example in one dimension). In three dimensions, exchange and correlation (both non-local and averaged, as here) play a crucial role in the formation of charge density waves<sup>13</sup>; in this respect, we note that in solid hydrogen near ninefold compression local electronic charge densities are both exceedingly high and rapidly varying. A reduction in the band-structure energy favouring the charge density wave state is balanced by an energy penalty originating with the short-range repulsion term in the effective intermolecular interaction. The strong dependence of the band structure on molecular orientation<sup>10</sup> offers an explanation of why charge density waves are favourable in some structures but not in others (for example,  $Pca2_1$  near ninefold compression; this is also reflected in the dependence of  $f$  on structure in the previously discussed classical model of spontaneous polarization).

Effects accompanying proton or deuteron motion are significant. As a function of pressure, the effect of which is generally to increase rotational barriers, hydrogen can pass from fully rotational to hindered (librational)<sup>14,15</sup>. The molecular librations act to weaken the collective polarizing field expected for fixed molecules. The reduction is necessarily larger in hydrogen than it is in deuterium because, although the rotational barriers for each are similar (they are largely electronic in origin), the moments of inertia differ by a factor of two. It follows that even for the ground state, the excursions of the axes of the deuterium molecule are considerably smaller than those of the hydrogen molecule, and the average local field at a deuterium molecule is therefore larger.

For both hydrogen and deuterium there is a discontinuous decrease of  $\sim 100 \text{ cm}^{-1}$  in the vibron as the phase II (hindered)–phase III (hindered and polarized) boundary is crossed at low temperatures at  $\sim 1.5$  million atmospheres. Because of the larger mass of deuterium, this corresponds to a fractionally larger change in its vibron and, viewed in terms of the harmonic approximation, the inference is that the relevant force constant in deuterium decreases more across the phase II–phase III boundary than it does in hydrogen. The explanation is now evident: the reduction of the collective effect, or polarizing field, due to librational motion must be smaller in deuterium than in hydrogen. The strong temperature dependence of the vibron discontinuity also implies that the underlying physical mechanisms must be of a fundamentally collective nature, and the weakening of the discontinuity as temperature is increased again indicates the reduction of the collective effect as librational amplitude increases. Finally, the large amplitude librational motion becomes increasingly incoherent



as the phase III–phase I boundary is approached. Subsequent crossing into phase I is accompanied by transition to full rotational behaviour of the pairs, an almost complete lack of coherence, and destruction of the essential self-consistent and coherent collective structure implicit in a charge density wave.

To compare the measured change in the vibron wavenumber  $\Delta\nu$  with that expected from the transition to a spontaneously polarized state, we have used the frozen-phonon method<sup>16</sup>, taking the structure  $Cmc2_1$  (Fig. 1A) which has the lowest energy of the structures considered here. At  $r_s = 1.47$  (9.6-fold compression), we find  $\Delta\nu = 170 \text{ cm}^{-1}$  for hydrogen, and  $\Delta\nu = 120 \text{ cm}^{-1}$  for deuterium. To determine the effect of librational motion we appeal to the experimentally determined result that  $\Delta\nu$  is directly proportional to the molecular polarization, or dipole moment<sup>12</sup>. We replace  $\mathbf{d}$  with  $\langle \mathbf{d} \rangle$ , estimating the librational amplitude using the calculation by Mazin and Cohen<sup>17</sup>. For an anisotropically polarizable linear molecule, with polarizability tensor components (in a frame in which the molecular axis lies along  $z$ )  $\alpha_{xx} = \alpha_{yy} = \alpha_{\perp}$ ,  $\alpha_{zz} = \alpha_{\parallel}$  (all other components vanishing), which librates in the field  $\mathbf{E} = E_0 \mathbf{z}$ , we have  $\langle \mathbf{d} \rangle = E_0 \mathbf{z} (\alpha_{\parallel} \langle \cos^2 \theta \rangle + \alpha_{\perp} \langle \sin^2 \theta \rangle)$ . Here  $\theta$  is the angle the molecular axis makes with the direction of the field  $\mathbf{E}$ . In the absence of librational motion this reduces to  $\mathbf{d}_0 = \alpha_{\parallel} E_0 \mathbf{z}$ , so that  $\mathbf{d}$  and thus also  $\Delta\nu$  are reduced by a factor  $\langle \mathbf{d} \rangle / \mathbf{d}_0 = 1 - \langle \sin^2 \theta \rangle (1 - \alpha_{\perp} / \alpha_{\parallel})$ . This leads to downward shifts  $\Delta\nu \approx 138 \text{ cm}^{-1}$  for hydrogen, and  $\Delta\nu \approx 103 \text{ cm}^{-1}$  for deuterium, results which are certainly approximate but nevertheless agree quite well with experiment.

Our results are for hydrogen molecules originally located at the sites of a hexagonal close-packed lattice but undergoing excursions from exact sites. But it is entirely possible that another structure, perhaps with a larger unit cell, may ultimately be favoured, yet which also develops spontaneous polarization; indeed, as we have verified, the off-site behaviour accompanied by spontaneous polarization is not limited to a single structure. The displacement of the molecules shown in Fig. 1 is equivalent to a sliding of alternate layers towards a base-centred orthorhombic structure ( $Cmca$  for the orientations in Fig. 1A)<sup>18</sup>, but with the displacement stopping well short of that configuration. This suggests the possibility that the layering transition discussed in ref. 18 may occur quite gradually.

An issue raised by this physical picture is the ultimate development of the spontaneous polarization at even higher densities. It has been suggested, for example, that molecular hydrogen may become fully ionic with sufficient compression<sup>19</sup>. For densities higher than those corresponding to  $r_s \approx 1.37$ , we find that the charge asymmetry begins to slowly decrease, a trend which is here established only for static protons and for fixed structure, so that further structural transitions are not ruled out. However, the system does not seem to be progressing towards a fully ionic state.

We suggest that phonons other than the vibron may provide further experimental evidence in support of the broken electronic symmetry that we propose, as optical phonons (other than the vibron) and librons must also become strongly infrared-active as the electronic polarization increases. This may provide a very simple explanation for the recent observation of a high-frequency ( $1,600 \text{ cm}^{-1}$ ) infrared-active phonon<sup>20</sup> in solid hydrogen at pressures in excess of 1.5 million atmospheres, and it suggests that detailed infrared studies at these and lower frequencies may offer considerable insight into the structural issues. With respect to other experimental signatures, a straightforward group-theory analysis of the structures considered here indicates the following: first,  $C2/m$  has two vibrons, one infrared-active and one Raman-active, in agreement with experiment. Second,  $Cmc2_1$  has two vibrons, and both are infrared and Raman-active. Third,  $Pca2_1$  has four vibrons; three of these are infrared-active whereas all four are Raman-active<sup>21</sup>. Of course, X-ray or neutron diffraction studies at a level sufficient to determine not only unit-cell parameters but also the contents of the unit cell would provide the ultimate test of the structure.  $\square$

Received 3 March; accepted 11 June 1997.

- Wigner, E. & Huntington, H. B. On the possibility of a metallic modification of hydrogen. *J. Chem. Phys.* **3**, 764–770 (1935).
- Hemley, R. J. & Mao, H. K. in *Elementary Processes in Dense Plasmas* (eds Ichimaru, S. & Ogata, S.) 271–282 (Addison Wesley, Reading, MA, 1995).
- Hemley, R. J. & Mao, H. K. Phase transition in solid molecular hydrogen at ultrahigh pressures. *Phys. Rev. Lett.* **61**, 857–860 (1988).
- Lorenzana, H. E., Silvera, I. F. & Goettel, K. A. Evidence for a structural phase transition in solid hydrogen at megabar pressures. *Phys. Rev. Lett.* **63**, 2080–2083 (1989).
- Hanfland, M., Hemley, R. J. & Mao, H. K. Novel infrared vibron absorption of solid hydrogen at megabar pressures. *Phys. Rev. Lett.* **70**, 3760–3763 (1993).
- Luttinger, J. M. & Tisza, L. Theory of dipole interactions in crystals. *Phys. Rev.* **70**, 954–964 (1946).
- Herzfeld, K. F. On atomic properties which make an element a metal. *Phys. Rev.* **29**, 701–705 (1927).
- Goldhammer, D. A. *Dispersion und Absorption des Lichtes* (Teubner, Leipzig, 1911).
- Teter, M. P., Payne, M. C. & Allan, D. C. Solution of Schrödinger's equation for large systems. *Phys. Rev. B* **40**, 12255–12263 (1989).
- Kaxiras, E. & Broughton, J. Energetics of ordered structures in molecular hydrogen. *Europhys. Lett.* **17**, 151–155 (1992).
- Nagara, H. & Nakamura, T. Stable phases of solid hydrogen at megabar pressures and at zero temperature. *Phys. Rev. Lett.* **68**, 2468–2471 (1992).
- Mazin, I. I. et al. Quantum and classical orientational ordering in solid hydrogen. *Phys. Rev. Lett.* **78**, 1066–1069 (1997).
- Overhauser, A. W. Exchange and correlation instabilities of simple metals. *Phys. Rev.* **167**, 691–698 (1968).
- Pauling, L. The rotational motion of molecules in crystals. *Phys. Rev.* **36**, 430–443 (1930).
- Stern, T. E. The symmetric spherical oscillator and the rotational motion of homopolar molecules in crystals. *Proc. R. Soc. Lond. A* **130**, 551–557 (1931).
- Yin, M. T. & Cohen, M. L. Microscopic theory of the phase transformation and lattice dynamics of Si. *Phys. Rev. Lett.* **45**, 1004–1007 (1980).
- Mazin, I. I. & Cohen, R. E. Insulator-metal transition in solid hydrogen: Implications of electronic-structure calculations for recent experiments. *Phys. Rev. B* **52**, R8597–R8600 (1995).
- Edwards, B., Ashcroft, N. W. & Lenosky, T. Layering transitions and the structure of dense hydrogen. *Europhys. Lett.* **34**, 519–524 (1996).
- Baranowski, B. A hypothesis concerning the low temperature phase transition in solid hydrogen and deuterium at about 150 GPa. *Polish J. Chem.* **66**, 1737–1740 (1992).
- Chen, N., Sterer, E. & Silvera, I. F. Extended infrared studies of high pressure hydrogen. *Phys. Rev. Lett.* **76**, 1663–1666 (1996).
- Cui, L., Chen, N. H. & Silvera, I. F. Excitations, order parameters, and phase diagram of solid deuterium at megabar pressures. *Phys. Rev. B* **51**, 14987–14997 (1995).

**Acknowledgements.** We thank M. Teter (Cornell Univ. and Corning Inc.) and D. Allan (Corning Inc.) for use of the Corning LDA program and for many discussions. We also thank R. J. Hemley for discussions of his experimental results.

Correspondence should be addressed to N.W.A. (e-mail: nwa@msc.cornell.edu).

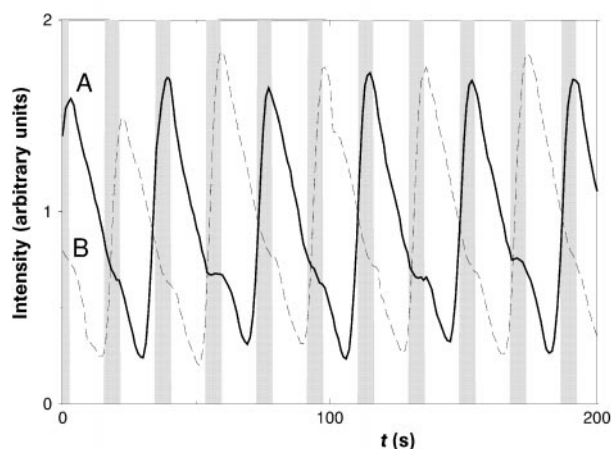
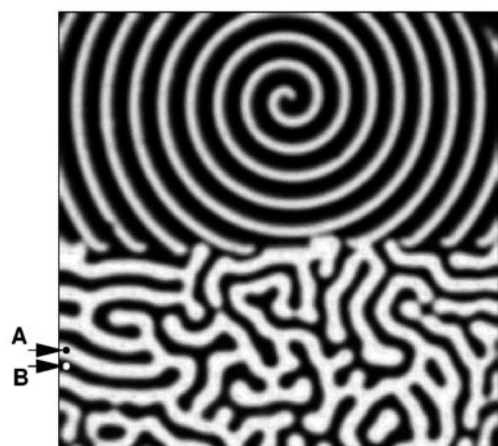
## Resonant pattern formation in a chemical system

Valery Petrov, Qi Ouyang & Harry L. Swinney

Center for Nonlinear Dynamics and Department of Physics, University of Texas, Austin, Texas 78712, USA

A periodic force applied to a nonlinear pendulum can cause the pendulum to become entrained at a frequency that is rationally related to the applied frequency, a phenomenon known as frequency-locking<sup>1</sup>. A recent theoretical analysis showed that an array of coupled nonlinear oscillators can exhibit spatial reorganization when subjected to external periodic forcing<sup>2</sup>. We present here experimental evidence that reaction–diffusion processes, which govern pattern evolution and selection in many chemical and biological systems<sup>3</sup>, can also exhibit frequency-locking phenomena. For example, periodic optical forcing of the light-sensitive Belousov–Zhabotinsky (BZ) reaction transforms a rotating spiral wave<sup>4</sup> to a labyrinthine standing-wave pattern (Fig. 1). As the forcing frequency is varied, we observe a sequence of frequency-locked regimes, analogous to the frequency-locked ‘tongues’ of a driven nonlinear pendulum, except that in the reactor different frequencies correspond to different spatial patterns. Resonant interactions leading to standing-wave patterns have not been observed previously in chemical or biological media, but periodic forcing (such as circadian rhythm) is abundant in nature and may lead to similar pattern-forming phenomena.

We examine the effect of external forcing on spiral wave patterns in a quasi-two-dimensional reaction–diffusion system with a light-sensitive form of the BZ reaction<sup>5,6</sup>. The reaction medium is a



thin membrane that is sandwiched between two reservoirs of reagents<sup>7</sup>. For the conditions of our experiment (see Fig. 1 legend), the reactor can oscillate homogeneously with the natural frequency  $f_0 = 0.028$  Hz; thus the reaction–diffusion medium can be thought of as a two-dimensional array of nonlinear oscillators coupled by diffusion. The natural frequency is measured by exposing the reaction to bright light for one minute; this resets each point in the reactor to the same initial conditions, and after the light is removed, the medium oscillates homogeneously for several cycles. However, the presence of boundaries and small imperfections in the reactor always leads to the appearance of rotating spirals. The rotation frequency of the spirals ( $f_{\text{spiral}} = 0.045$  Hz) is larger than that of the homogeneous oscillations, and the pattern evolves into a single spiral (top half of Fig. 1), owing to the dominance of the fastest pacemaker in the reactor<sup>8</sup>.

The reaction is perturbed at frequency  $f_p$  with pulses of spatially uniform light from a video projector. The absorption of light by ruthenium molecules causes a temporary shift of the chemical kinetics<sup>6</sup>. The periodic perturbations can destroy the spiral wave pattern and lead to a labyrinthine standing wave pattern, as shown in the (illuminated) lower half of Fig. 1. Black and white domains in the labyrinthine pattern correspond respectively to high and low concentrations of a ruthenium(II) complex. These domains oscillate with opposite phase and are separated by nodal lines that are time-independent. Figure 2 demonstrates that each domain oscillates in synchrony with every other perturbation pulse ( $f_p/f_0 = 2$ )—a 2:1 resonance. The observed effect of forcing on our oscillatory BZ system is different from that found previously in an excitable BZ

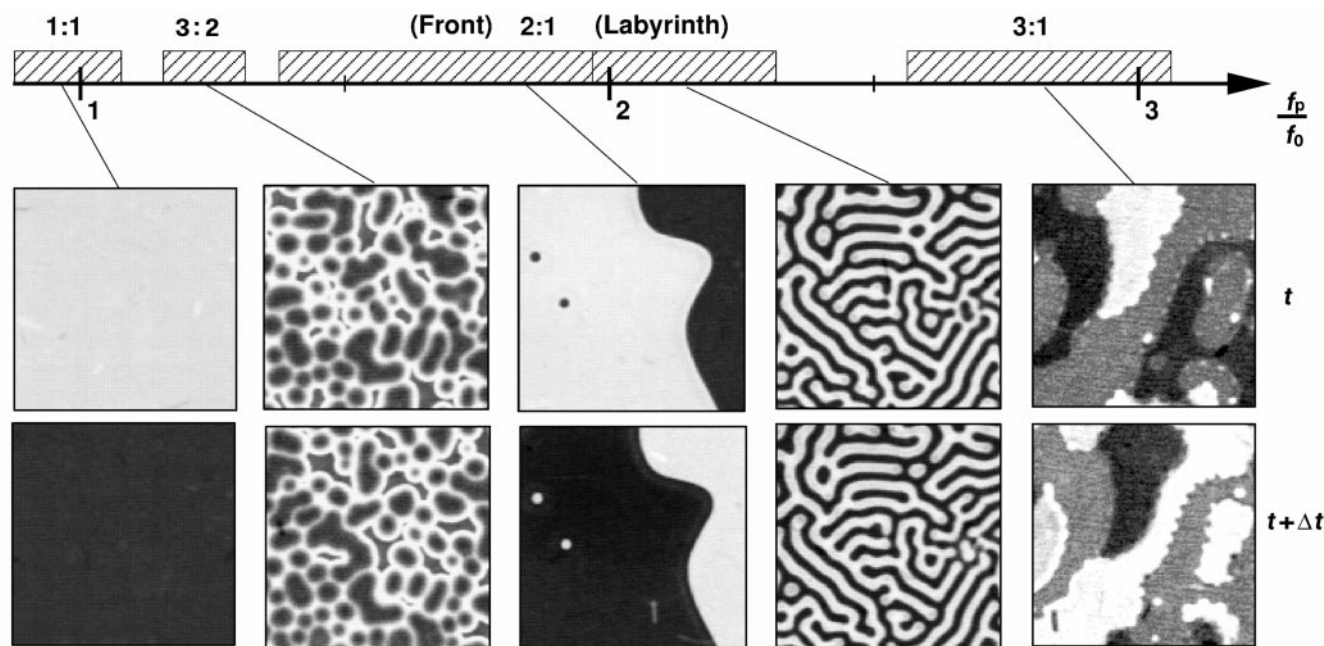
**Figure 1** A labyrinthine standing-wave pattern forms in the lower half of the reactor, which is illuminated with light pulsed at twice the natural frequency of the reaction ( $f_p/f_0 = 2$ ), while a spiral wave forms in the upper half of the same reactor, which is in the dark. The reaction occurs in a 0.4-mm-thick porous membrane disk that is placed between two reservoirs (I and II) which are continuously refreshed with BZ reagents. The region shown is a  $13 \times 13$  mm section of the 25-mm-diameter reactor. The patterns in the membrane are determined from light absorption measurements at 450 nm. In the perturbed region, light of  $0.2 \text{ W m}^{-2}$  intensity in the spectral range 430–470 nm is periodically switched on for 6 s and then off for 13 s. Spatially uniform illumination is achieved using a video projector (Sanyo PLC-220N) calibrated by measuring light reflected from a scatter plate and then adjusting each pixel to have the same intensity. The concentrations of the chemicals in reservoirs I and II are: [malonic acid]<sub>I</sub> = 0.22 M,  $[\text{BrO}_3^-]_{\text{I}} = 0.046$  M,  $[\text{Br}^-]_{\text{I}} = 0.2$  M,  $[\text{H}_2\text{SO}_4]_{\text{I}} = 0.8$  M,  $[\text{Ru}(\text{bpy})_3^{2+}]_{\text{I}} = 1.0$  mM,  $[\text{H}_2\text{SO}_4]_{\text{II}} = 0.8$  M,  $[\text{BrO}_3^-]_{\text{II}} = 0.184$  M (here bpy indicates 2,2'-bipyridine). The volume of each reservoir is 10 ml. The flow rate in reservoir I is  $20 \text{ ml h}^{-1}$ ; in reservoir II,  $5 \text{ ml h}^{-1}$ . The system is maintained at a temperature of  $23^\circ\text{C}$ .

**Figure 2** Time series of pattern intensity at the points A (solid line) and B (dashed line) in Fig. 1. The grey-shaded stripes indicate when light perturbations are applied to the reactor. The oscillations at A and B are opposite in phase and synchronized at one-half of the forcing frequency.

reaction in a Petri dish<sup>9</sup>. Unlike the oscillatory medium, the excitable BZ reaction cannot oscillate homogeneously but can form rotating spirals. Periodic forcing of excitable spirals has been found to lead to a modulation of the rotation frequency and a slow drift of the spiral centre, but not to standing waves<sup>9</sup>.

When we vary the frequency of perturbations, we observe a sequence of resonance patterns, each persisting for a range of perturbation frequency. Figure 3 illustrates the patterns for frequency-locked ratios  $f_p/f_0 = 1, 3/2, 2$  and  $3$ . In the 1:1 regime, the entire reactor synchronizes with the perturbation, oscillating between uniform dark and light. When  $f_p$  is increased, the system is no longer able to synchronize with the perturbation; instead, bubble-shaped structures appear (Fig. 3). The bubbles appear and disappear in a seemingly random fashion, but the temporal spectrum at any spatial point in the pattern has well defined peaks at multiples of  $f_p/3$ , indicating that this is a 3:2 resonance.

Two types of 2:1 resonances are observed: for  $f_p/f_0 < 1.9 \pm 0.1$ , the pattern consists of a stable stationary front separating two oscillating, spatially uniform domains. The shape of the domains does not change; it is determined by the concentration profiles in the beginning of experiment. We also observe localized spots that coexist with the stationary front (Fig. 3); similar localized structures have been found in models of bistable chemical reactions<sup>10</sup> and coupled map lattices<sup>11</sup>. For  $f_p/f_0 > 1.9 \pm 0.1$ , the previously described labyrinthine pattern forms. Once this pattern has grown to fill the entire perturbed region of the reactor, the nodal lines between the two types of domains become stationary. The transition from simple fronts to labyrinthine standing wave is similar to the



**Figure 3** Bifurcation diagram showing different frequency-locked regimes observed as a function of  $f_p/f_0$  where  $f_p$  is the perturbation frequency and  $f_0$  is the natural frequency of the system. Patterns are shown in pairs, one above the other, at times separated by  $\Delta t = 1/f_p$ , except for the 1:1 resonance where  $\Delta t = 1/2f_p$ . Striped boxes on the horizontal axis mark perturbation frequency ranges with the same frequency-locking ratio. The frequency of the pattern

oscillations is determined by analysing time series of single-point measurements similar to that shown in Fig. 2 for the 2:1 resonance. The duration of the perturbation light pulses is fixed at 6 s for all measurements; at  $f_p/f_0 = 1$  the period of the forcing is 36 s, and at  $f_p/f_0 = 3$  the period of the forcing is 12 s. Images are available at <http://chaos.ph.utexas.edu/~lera/forcing.html>

transition from planar fronts to a stationary labyrinthine pattern observed recently<sup>12</sup> in an intrinsically bistable (unforced) reaction. This indicates that in the 2:1 resonance, the BZ reaction behaves as a bistable system with the two states corresponding to the two possible oscillatory phases.

The pattern in the 3:1 resonance regime has three types of domains which oscillate with phases differing by  $2\pi/3$ . After three periods of perturbation, the original pattern is recovered, except for a very slow drift of the domain boundaries. The observed boundary drift supports a theoretical conjecture that such behaviour is a fundamental property of interfaces in tristable systems<sup>13</sup>.

The bifurcation diagram in Fig. 3 has gaps where no resonance was evident. The temporal behaviour in these regions may be quasi-periodic rather than locked, or some of these regimes may be locked at frequency ratios  $m/n$  where  $m$  and  $n$  are too large to be determined in the present experiments, or the fluctuations of the experimental conditions may have shifted the system between adjacent regions. Experiments with a better signal-to-noise ratio are needed to distinguish between these and other possible explanations for the gap regions.

The sequence of resonance regimes observed in the present study is analogous to the behaviour of strongly forced low-dimensional nonlinear dynamical systems<sup>1</sup>, where, as the perturbation frequency is varied, the system can exhibit a sequence of regimes with frequency locking at ratios  $m/n$  (where  $m$  and  $n$  are integers). Unlike the low-dimensional systems that can be entrained by very small perturbations near resonance, the observed standing-wave patterns appear only for a forcing intensity above some critical value. At lower intensity, we observed rotating spirals and other non-stationary patterns.

A transition from travelling to standing waves has been observed previously in several physical systems in the 2:1 locking regime<sup>14</sup>. Our experiments reveal two types of patterns with 2:1 locking. Moreover, we find other frequency-locked regimes which have different types of spatiotemporal behaviour. Patterns similar to

some of those that we have observed have been obtained in a model of a forced weakly nonlinear medium described by a complex Ginzburg–Landau equation<sup>15</sup>. This similarity in the phenomena for a general model and for our experimental system suggests that resonant pattern formation may arise in a variety of chemical, biological and other extended systems where internal oscillatory dynamics are coupled with external periodic forcing. In three dimensions, where spirals become scroll waves<sup>16</sup>, periodic forcing may lead to yet another type of pattern. □

Received 11 March; accepted 17 June 1997.

- Thompson, J. M. T. & Stewart, H. B. *Nonlinear Dynamics and Chaos* (Wiley, Chichester, 1986).
- Coullet, P. & Emilion, K. Strong resonances of spatially distributed oscillators: a laboratory to study patterns and defects. *Physica D* **61**, 119–131 (1992).
- Murray, J. D. *Mathematical Biology* (Springer, Berlin, 1989).
- Winfree, A. T. Spiral waves of chemical activity. *Science* **175**, 634–636 (1972).
- Gaspar, V., Bazsa, G. & Beck, M. T. The influence of visible light on the Belousov–Zhabotinski oscillating reactions applying different catalysts. *Z. Phys. Chem. (Leipzig)* **264**, 43–48 (1983).
- Kuhnert, L. A new optical photochemical memory device in a light-sensitive chemical active medium. *Nature* **319**, 393–394 (1986).
- Ouyang, Q. & Swinney, H. L. Transition to chemical turbulence. *Chaos* **1**, 411–420 (1991).
- Krinsky, V. I. & Agladze, K. I. Interaction of rotating waves in an active chemical medium. *Physica D* **8**, 50–56 (1983).
- Steinbock, O., Zykov, V. & Müller, S. Control of spiral-wave dynamics in active media by periodic modulation of excitability. *Nature* **366**, 322–324 (1993).
- Hagberg, A. & Meron, E. Complex patterns in reaction-diffusion systems: A tale of two front instabilities. *Chaos* **4**, 477–484 (1994).
- Kapral, R., Livi, R., Oppo, G.-L. & Politi, A. Dynamics of complex interfaces. *Phys. Rev. E* **49**, 2009–2022 (1994).
- Lee, K. J., McCormick, W. D., Ouyang, Q. & Swinney, H. L. Pattern formation by interacting chemical fronts. *Science* **261**, 192–194 (1993).
- Bennett, C. H., Grinstein, G., He, Y., Jayaprakash, C. & Mukamel, D. Stability of temporally periodic states of classical many-body systems. *Phys. Rev. A* **41**, 1932–1935 (1990).
- Rehberg, I., Rasenat, S., Fineberg, J., de la Torre Juez, M. & Steinberg, V. Temporal modulation of travelling waves. *Phys. Rev. Lett* **61**, 2449–2452 (1988).
- Walgraef, D. *Spatio-Temporal Pattern Formation* (Springer, New York, 1997).
- Winfree, A. T. Scroll-shaped waves of chemical activity in three dimensions. *Science* **181**, 937–939 (1973).

**Acknowledgements.** We thank A. Lee for discussions and G. Li for help in conducting the experiments. This work was supported by the US Department of Energy Office of Basic Energy Sciences and the Robert A. Welch Foundation.

Correspondence should be addressed to H.L.S. (email: swinney@chaos.ph.utexas.edu).

# Plasticity and avalanche behaviour in microfracturing phenomena

Stefano Zapperi<sup>\*</sup>, Alessandro Vespignani<sup>†</sup>  
& H. Eugene Stanley<sup>\*</sup>

<sup>\*</sup> Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

<sup>†</sup> Instituut-Lorentz, University of Leiden, PO Box 9506, 2300 RA, Leiden, The Netherlands

Inhomogeneous materials, such as plaster or concrete, subjected to an external elastic stress display sudden movements owing to the formation and propagation of microfractures. Studies of acoustic emission from these systems reveal power-law behaviour<sup>1</sup>. Similar behaviour in damage propagation has also been seen in acoustic emission resulting from volcanic activity<sup>2</sup> and hydrogen precipitation in niobium<sup>3</sup>. It has been suggested that the underlying fracture dynamics in these systems might display self-organized criticality<sup>4</sup>, implying that long-ranged correlations between fracture events lead to a scale-free cascade of 'avalanches'. A hierarchy of avalanche events is also observed in a wide range of other systems, such as the dynamics of random magnets<sup>5</sup> and high-temperature superconductors<sup>6</sup> in magnetic fields, lung inflation<sup>7</sup> and seismic behaviour characterized by the Gutenberg–Richter law<sup>8</sup>. The applicability of self-organized criticality to microfracturing has been questioned<sup>9,10</sup>, however, as power laws alone are not unequivocal evidence for it. Here we present a scalar model of microfracturing which generates power-law behaviour in properties related to acoustic emission, and a scale-free hierarchy of avalanches characteristic of self-organized criticality. The geometric structure of the fracture surfaces agrees with that seen experimentally. We find that the critical steady state exhibits plastic macroscopic behaviour, which is commonly observed in real materials.

Quasi-static models of fracture propagation have been extensively studied in the past<sup>11</sup>. Such analyses have been mainly focused on the geometrical properties of the macroscopic crack, although some dynamical effects have also been studied<sup>12,13</sup>. Recently, avalanches and other dynamical properties have been investigated in a model for hydraulic fracturing<sup>14</sup>. However, the amplitude distribu-

tion associated with the acoustic emission signal did not reveal any scaling behaviour.

The mesoscopic description of an elastic disordered medium is usually obtained by discretizing macroscopic elastic equations. In the theory of linear elasticity<sup>15</sup>, these equations relate the stress tensor  $s_{\alpha\beta}$  to the strain tensor  $\epsilon_{\gamma\delta}$  via the Hooke tensor  $C_{\alpha\beta\gamma\delta}$ :

$$s_{\alpha\beta} = \sum_{\gamma,\delta} C_{\alpha\beta\gamma\delta} \epsilon_{\gamma\delta} \quad (1)$$

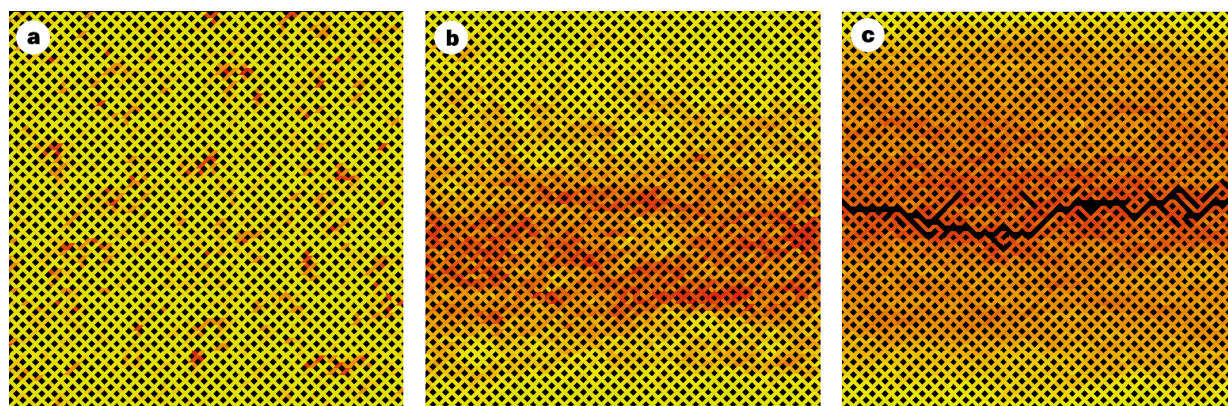
In highly inhomogeneous materials, like concrete, linear elasticity breaks down owing to the formation and the propagation of microcracks. It is still possible to describe the system using equation (1) by using an effective Hooke tensor  $\tilde{C}$ . The 'damage'  $D$  is defined through the relation between  $C$  and  $\tilde{C}$  (ref. 16),

$$\tilde{C} = (1 - D)C \quad (2)$$

where the tensor indices have been omitted for simplicity. The full tensorial formalism can be quite complex to handle numerically. Fortunately, many essential features of fracture phenomenology can be described using scalar models<sup>11,17,18</sup>. There is a formal analogy between the scalar elasticity problem and an equivalent electrical problem; one identifies the current  $I$  with the stress, the voltage  $V$  with the strain, and the conductivity  $\sigma$  with the Hooke tensor: we will use this electrical analogy.

To study the system at a mesoscopic scale, we discretize the problem by considering a resistor network<sup>17</sup> on a tilted square lattice. We introduce the disorder, due to the inhomogeneities present in the material, by assigning a random failure threshold  $I_c$  to each resistor, where  $I_c$  is drawn from a uniform distribution in the interval  $[0,1]$ . In the classic 'fuse' model<sup>17,19</sup>, if the current flowing in a resistor exceeds the failure threshold, the bond is removed from the lattice (that is, the electrical conductivity drops to zero). In this way the system develops a macroscopic crack and eventually the lattice breaks apart. We adopt a different breaking rule: when the current in a bond exceeds the threshold, we impose permanent 'damage' to the bond by decreasing the conductivity of the bond by a factor  $a \equiv (1 - D)$ . Describing the local damage with a continuous parameter corresponds to studying the system at a scale larger than the microscopic crack scale, but smaller than the homogeneous macroscopic scale.

After a bond failure we choose at random a new threshold for the bond in order to model the microscopic rearrangements in the material. The model in this form is quite slow to simulate numerically. A big reduction in the simulation time can be obtained by also



**Figure 1** The damage for a system of size  $L = 64$ . **a**, In the early stage, after 500 avalanches; **b**, in the steady state, after 1,000 avalanches. **c**, The final crack, which develops in the highly damaged region. Damaged parts are shown in red and undamaged parts in yellow; absent bonds (the crack) are shown in black.



changing the thresholds of the bonds that neighbour a damaged bond. This local rearrangement does not change the properties of the model but makes it easier to obtain statistics of sufficiently high quality.

The simulation proceeds as follows. We apply an external voltage difference between two opposite edges of the lattice, while periodic boundary conditions are imposed on the other direction. We compute the currents in each bond by solving numerically the Kirchhoff equations, using a multigrid<sup>20</sup> relaxation algorithm with precision  $\epsilon = 10^{-12}$ . The voltage is then slowly increased until the current in some bond reaches its threshold. The bond is damaged and the disorder is changed according to the rules specified above. We then compute new currents and repeat the process until no unstable bonds are present. Owing to the long-range elastic interactions and the 'redistribution' of the disorder, a single bond failure can be followed by an avalanche of additional failures.

We study the system in the limit of 'slow driving': the timescale over which fractures form and propagate is much faster than the timescale of the external driving. This is a common characteristic of systems that display self-organized criticality (SOC). In fact, for SOC systems the control parameter is always related to the ratio between two timescales. In this situation, however, the existence of a timescale separation makes the system very close to the critical point for a wide range of internal parameters.

We start the simulation with undamaged material: all the conductivities are set equal to one. In the early stage of the process only small rearrangements take place. This is also evident by observing the structure of the microfractures in the system (Fig. 1a): the damage is homogeneously scattered throughout the system. Increasing the voltage leads to a corresponding increase of the total current flowing in the system: macroscopically the material behaves elastically. Deviations from linear elasticity start to appear as the activity increases. Eventually, the system reaches a steady-state which is macroscopically 'plastic' (Fig. 2), in the sense that the increase of the voltage is balanced by the damage in such a way that the current is kept approximately constant. In this state the activity is highly fluctuating and avalanches of all sizes occur. Plastic behaviour is not uncommon in stressed synthetic materials<sup>21</sup>.

The damage in the steady state tends to be organized in linear bands (Fig. 1b), arising from the coalescence of the underlying microfractures. A similar geometrical structure has been observed in microfracturing experiments on concrete<sup>22,23</sup>. We note that a plastic steady state characterized by highly fluctuating activity has been recently obtained in molecular dynamic studies of granular solids<sup>24</sup> and foams<sup>25</sup> under shear.

In the plastic steady state, we investigate the statistical properties of rupture sequences (Fig. 3a) in time and magnitude during fracturing. We define the avalanche size  $s$  as the number of bonds damaged for a given voltage increment, and find that the avalanche size probability distribution function  $P(s)$  exhibits a power-law behaviour

$$P(s) \sim s^{-\tau} \quad (3)$$

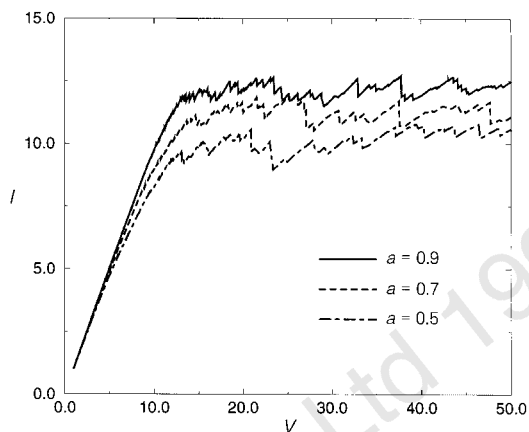
where  $\tau = 1.19 \pm 0.01$ . By computing the distribution for different lattice sizes, we observe that the cut-off scales with the system size (Fig. 3b). This is the fingerprint of a scale-free activity; an absence of characteristic lengths in the system.

We next compute the distribution of the time duration  $T$  of each avalanche, obtaining a power-law decay

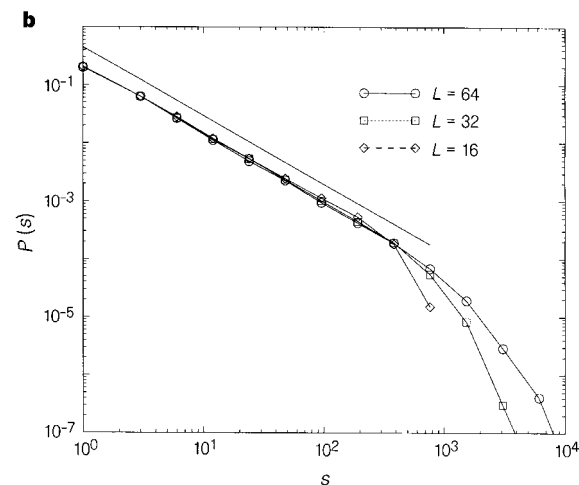
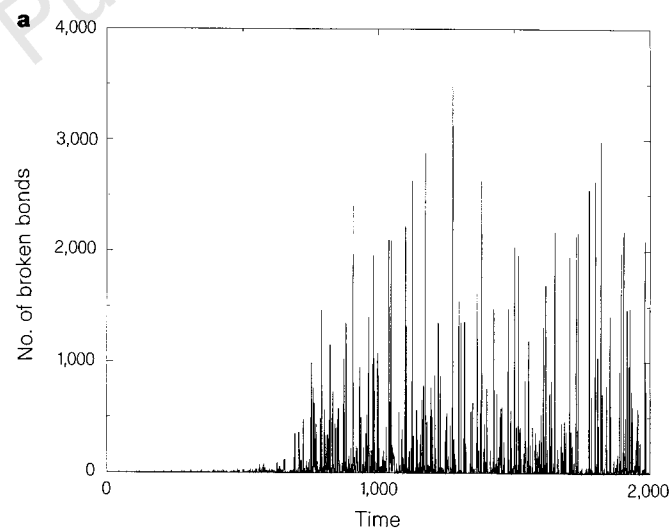
$$P(T) \sim T^{-\alpha} \quad (4)$$

where  $\alpha = 1.30 \pm 0.01$ . Again we find that the cut-off scales with the system size.

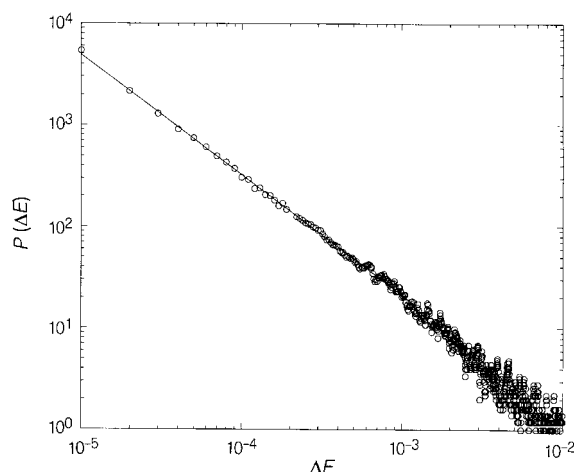
We also study the distribution of energy bursts, which is directly related to the acoustic-emission signal recorded experimentally. In this case also we find a power-law distribution if the energy is rescaled by the energy of the unbroken lattice, as pointed out in



**Figure 2** The  $I$ - $V$  characteristics of the present model for different values of the parameter  $a$  (see text). Note the crossover at  $V \approx 12$  between linear elasticity and plasticity.



**Figure 3 a**, The number of broken bonds  $s$  as a function of time. **b**, The avalanche size distribution for different system sizes ( $L = 16, 32, 64$ ). The statistical analysis is performed only in the steady state for  $a = 0.9$ , averaging over  $2 \times 10^4$  avalanches.



**Figure 4** The distribution of energy bursts, related to the experimentally observed acoustic emission. We fit the power-law distribution with an exponent equal to  $1.2 \pm 0.1$ .

refs 26 and 27. The results for a lattice of size  $L = 64$  are shown in Fig. 4.

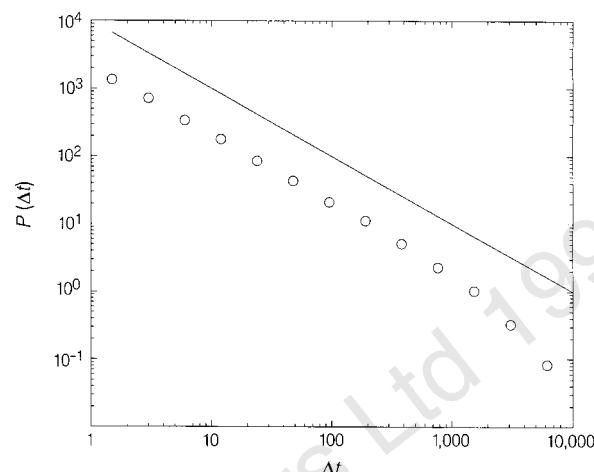
Another quantity of interest is the distribution of time intervals  $\Delta t$  between two avalanches in the steady state (Fig. 5). We find a power decay which is reminiscent of the Omori law in earthquake statistics<sup>28</sup>

$$P(\Delta t) \sim (\Delta t)^{-\gamma} \quad (5)$$

where  $\gamma \approx 1$ . The above behaviour provides another indication that the system is critical in the plastic steady state, which develops critical correlations and scaling behaviour through a self-organization process. The critical exponents seem to be independent of the changes in the damage parameter  $a$  within numerical uncertainties.

In our model the steady state could in principle last forever. However, in realistic situations, the damage description should fail at a given point. After a certain number of failure events, a bond will no longer respond elastically. This observation can be modelled in a natural way by allowing for only a finite number of failures per bond. In other words, when the conductivity of a bond reaches a given value  $\sigma_c$  owing to the repeated failures, the bond is no longer considered to be elastic and is removed from the lattice. The parameter  $\sigma_c$  adjusts the duration of the steady-state: for high values of  $\sigma_c$ , the steady state is never reached and fracture is brittle, whereas a low value of  $\sigma_c$  produces a plastic steady state. In both cases a macroscopic crack will eventually form. In Fig. 1c we observe that the crack develops in the region of high damage and has a rough structure. Owing to the limited system size, we were not able to calculate reliably the roughness exponent.

We note that to observe this behaviour the system must be driven at a constant voltage. We have also performed simulations in which the system is driven by imposing a constant current: in this case no steady state is observed and the system is driven to an instability corresponding to the critical current of the voltage-driven experiment. One can still fit the non-stationary distribution of avalanches with a power law but the system is clearly not SOC. With our model we can therefore reconcile the controversy<sup>9,10</sup> between SOC and 'sweeping of an instability'<sup>29</sup> as possible explanations of microfracturing experiments: either of the two phenomena can arise according to the driving conditions. It would be interesting to test this prediction experimentally by comparing avalanche signals in stress-driven and strain-driven experiments. In interpreting experimental results, it is important to keep in mind that SOC requires a 'stationary' signal. In fracture phenomena this can be obtained in the plastic regime. □



**Figure 5** The distribution of quiescent intervals  $\Delta t$  between two avalanches in the plastic state. Plotted for reference is a line with slope corresponding to  $\gamma = 1$  in equation (5).

Received 2 April; accepted 3 June 1997.

- Petri, A., Paparo, G., Vespignani, A., Alippi, A. & Costantini, M. Experimental evidence for critical dynamics in microfracturing processes. *Phys. Rev. Lett.* **73**, 3423–3426 (1994).
- Diodati, P., Marchesoni, F. & Piazza, S. Acoustic emission from volcanic rocks: an example of self-organized criticality. *Phys. Rev. Lett.* **67**, 2239–2242 (1991).
- Cannelli, G., Cantelli, R. & Cordero, F. Self-organized criticality of the fracture processes associated with hydrogen precipitation in niobium by acoustic emission. *Phys. Rev. Lett.* **70**, 3923–3926 (1993).
- Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: an explanation of  $1/f$  noise. *Phys. Rev. Lett.* **59**, 381–384 (1987).
- Cote, P. J. & Meisel, L. V. Self-organized criticality and the Barkhausen effect. *Phys. Rev. Lett.* **67**, 1334–1337 (1991).
- Field, S., Witt, J., Nori, F. & Ling, X. Superconducting vortex avalanches. *Phys. Rev. Lett.* **74**, 1206–1209 (1995).
- Suki, B., Barabási, A.-L., Hantos, Z., Peták, F. & Stanley, H. E. Avalanches and power-law behaviour in lung inflation. *Nature* **368**, 615–618 (1994).
- Gutenberg, B. & Richter, C. F. Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* **34**, 185–188 (1944).
- Sornette, D. Power laws without parameter tuning: an alternative to self-organized criticality. *Phys. Rev. Lett.* **72**, 2306 (1994).
- Cannelli, G., Cantelli, R. & Cordero, F. Reply to Sornette D., 'Power laws without parameter tuning: an alternative to self-organized criticality'. *Phys. Rev. Lett.* **72**, 2307 (1994).
- Herrmann, H. J. & Roux, S. (eds) *Statistical Models for the Fracture of Disordered Media* (North Holland, Amsterdam, 1990).
- Sornette, D. & Vanneste, C. Dynamics and memory effects in rupture of thermal fuse networks. *Phys. Rev. Lett.* **68**, 612–615 (1992).
- Miltenberger, P., Sornette, D. & Vanneste, C. Fault self-organization as optimal random paths selected by critical spatiotemporal dynamics of earthquakes. *Phys. Rev. Lett.* **71**, 3604–3607 (1993).
- Tzschichholz, F. & Herrmann, H. J. Simulations of pressure fluctuations and acoustic emission in hydraulic fracturing. *Phys. Rev. E* **51**, 1961–1970 (1995).
- Landau, L. D. & Lifschitz, E. M. *Theory of Elasticity* (Pergamon, New York, 1960).
- Wilshire, B. & Owen, D. R. J. *Engineering Approaches to High Temperature Design* (Pineridge, Swansea, UK, 1983).
- De Arcangelis, L., Redner, S. & Herrmann, H. J. A random fuse model for breaking processes. *J. Phys. (Paris)* **46**, L585–L590 (1985).
- Herrmann, H. J., Kertész, J. & de Arcangelis, L. Fractal shapes of deterministic cracks. *Europhys. Lett.* **10**, 514–519 (1991).
- De Arcangelis, L. & Herrmann, H. J. Scaling and multiscaling laws in random fuse networks. *Phys. Rev. B* **39**, 2678–2684 (1989).
- Press, W. H. & Teukolski, S. A. Multigrid methods for boundary value problems. *Comput. Phys.* **5**, 154–159 (1991).
- Chen, W. F. *Plasticity in Reinforced Concrete* (McGraw-Hill, New York, 1982).
- Stroeven, P. in *Interfaces and Cementitious Composites* (ed. Maso, J. C.) 187–196 (Spon, London, 1993).
- Stroeven, P. Some observations on microcracking in concrete subjected to various loading regimes. *Eng. Frac. Mech.* **35**, 775–782 (1990).
- Tillemans, H. J. & Herrmann, H. J. Simulating deformations of granular solids under shear. *Physica A* **217**, 261–288 (1995).
- Okuzono, T. & Kawasaki, K. Intermittent flow behavior of random foams: a computer experiment on foam rheology. *Phys. Rev. E* **51**, 1246–1253 (1995).
- Caldarelli, G., Di Tolla, F. & Petri, A. Self organization and annealed disorder in fracturing process. *Phys. Rev. Lett.* **77**, 2503–2506 (1996).
- Sahimi, M. & Arbabi, S. Scaling laws for fracture of heterogeneous materials and rock. *Phys. Rev. Lett.* **77**, 3689–3692 (1996).
- Omori, F. *J. Coll. Sci. Imper. Univ. Tokyo* **7**, 111 (1894).
- Sornette, D. Sweeping of an instability: an alternative to self-organized criticality to get powerlaws with parameter tuning. *J. Phys. I France* **4**, 209–221 (1994).

**Acknowledgements.** We thank G. Caldarelli, R. Cuerno, J. Kertész, H. J. Herrmann, K. B. Lauritsen, A. Petri, C. Rebbi and P. Stroeven for suggestions and discussions. The Center for Polymer Studies is supported by NSF.

Correspondence should be addressed to S.Z. (e-mail: zapperi@bu.edu).

# Ozone-rich transients in the upper equatorial Atlantic troposphere

K. Suhre\*, J.-P. Cammas\*, P. Nédélec\*, R. Rosset\*, A. Marengo\* & H. G. J. Smit†

\* Laboratoire d'Aérodynamique (UMR CNRS/UPS 5560), OMP, 14, Avenue Edouard Belin, 31400 Toulouse, France

† Research Centre Jülich, Institute for Chemistry of the Polluted Atmosphere (ICG-2), PO Box 1913, 52425 Jülich, Germany

High concentrations of ozone are found in the Earth's stratosphere, but strong stratification suppresses efficient exchange of this ozone-rich air with the underlying troposphere. Upward transport of tropospheric trace constituents occurs mainly through equatorial deep convective systems. In contrast, significant downward transport of ozone-rich stratospheric air is thought to take place only outside the tropics by exchange processes in upper-level fronts associated with strong distortions of the tropopause<sup>1</sup>. Ozone within the tropical troposphere is assumed to originate predominantly from ground-based emissions of ozone precursors, particularly from biomass burning<sup>2</sup>, rather than from a stratospheric source. Recent measurements of ozone in the upper troposphere in convective regions over the Pacific Ocean<sup>3</sup> indeed reveal near-zero concentrations. Here we present sharply contrasting observations: ozone-rich (100–500 parts per billion by volume) transients were frequently encountered by specially equipped commercial aircraft at a cruising altitude of 10–12 km (in the upper troposphere) in the vicinity of strong convective activity over the equatorial Atlantic Ocean. This strongly suggests that the input of stratospheric ozone into the troposphere can take place in the tropics. We suggest that this transport occurs either by direct downward movement of air masses or by quasi-isentropic transport from the extratropical stratosphere.

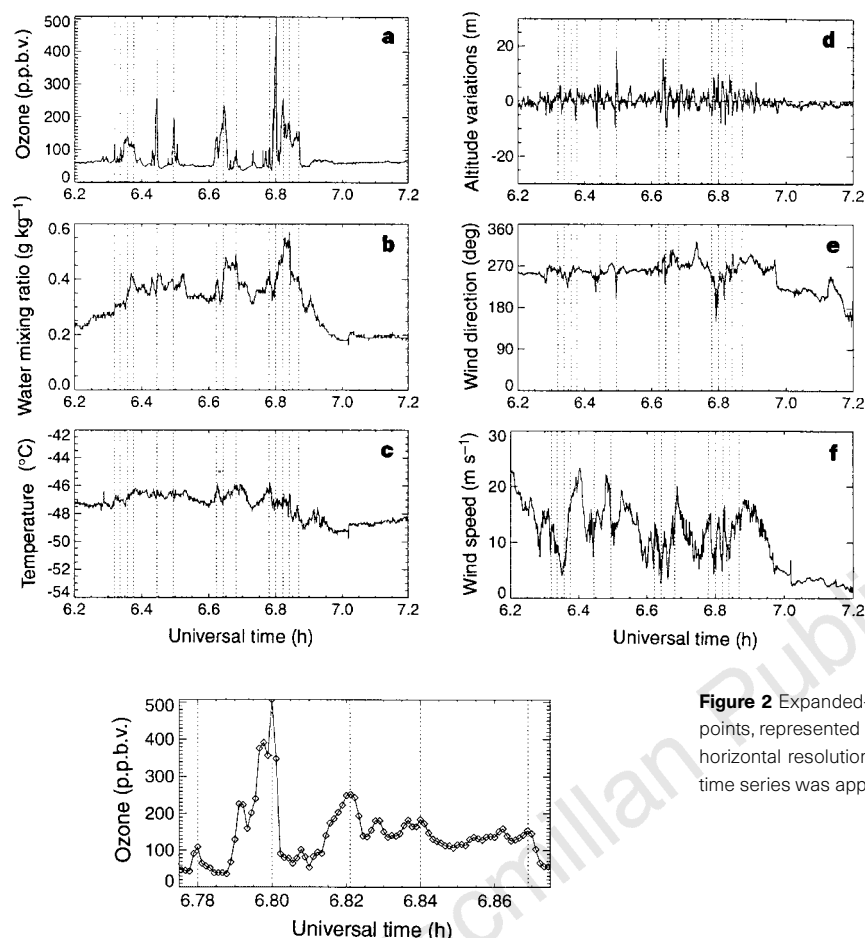
Five commercial Airbus 340 aircraft operated by Air France, Austrian Airlines and Lufthansa have been equipped with ozone, temperature and humidity instruments, as part of the 'Measurement of Ozone by Airbus In-Service Aircraft' (MOZAIC) project<sup>4,5</sup>. We report here an analysis of more than 100 flights across the Inter-Tropical Convergence Zone (ITCZ) over the tropical Atlantic Ocean between Europe and South America. During most of the Atlantic transect, ozone mixing ratios are well below 50 parts per billion by volume (p.p.b.v.) with only small fluctuations. But in one-third of the reported flights, uniquely within the 15°N–15°S equatorial latitude belt, we observe events where ozone mixing ratios suddenly increase for a short time period to over 100 p.p.b.v., and in some cases up to 500 p.p.b.v., at a horizontal scale of 5–80 km. A representative example is shown in Figs 1 and 2. At the same time, the aircraft enters a zone of turbulence of about the same horizontal extent, as the pressure–altitude records indicate a vertical displacement of the order of 10 m (Fig. 1d) and abrupt changes in wind speed and direction are observed (Fig. 1e, f). At the same time as these high-ozone events, water mixing ratios usually increase drastically (Fig. 1b) and temperatures slightly decrease (Fig. 1c), but not systematically for the latter. Relative humidity >100% is sometimes observed during MOZAIC flights and particularly during these high-ozone events. This is not ascribed to artefacts; the extra water is probably provided by the evaporation of water droplets due to adiabatic compression and associated heating which cause a substantial temperature increase in the humidity-sensor housing at cruising speeds of Mach 0.8. Thus, the high humidity values must be interpreted qualitatively as indicating the

presence of ice crystals during the sampling. The aircraft often deviates from its otherwise straight flight track during these high-ozone events, apparently to outflank an equatorial convective system. The presence of active deep convection has been identified on NOAA/AVHRR satellite images in most cases. Note that for safety reasons passenger aircraft never pass through the active part of tropical deep convective systems, but at a distance of 15–30 km around the convective cloud towers, thus passing under the anvil clouds and their precipitation streaks.

In 10% of all analysed flights across the ITCZ, an event with very high ozone mixing ratio (>150 p.p.b.v.) over a distance of some tens of kilometres has been observed, together with all the following: turbulence, decreasing temperature, rising humidity, changes in wind speed and direction, deviation of the aircraft trajectory from a straight line, and convective systems on NOAA/AVHRR images. In even more cases (30% of all flights), high ozone values (100–200 p.p.b.v.) have been found and correlated with some, but not always all, of the observations described above. These latter cases are likely to be the observation of residual perturbations some time after the event itself occurred, whereas the 10% of 'full' cases probably represent observations that took place shortly after the perturbations were caused. Here we consider only observations of 'full' cases made over the Atlantic Ocean, where MOZAIC data is most dense.

More than 5,000 MOZAIC flights in the upper troposphere (low-ozone regime) as well as in the lower stratosphere (high-ozone regime) show that the performance of the ozone analyser is not affected by the aircraft entering or leaving clouds (no significant ozone changes are observed), by turbulence (many cases of strong turbulence without change in ozone concentrations have been observed outside the tropics), or by icing of the ozone inlet (no significant changes of flow rates in the sampling line and ozone device have been observed). The rare erratic ozone values that are sometimes found consist in the worst case of only two successive records. On a large number of MOZAIC flights to South Africa and to Asia similar cases of high-ozone transients have been encountered when crossing the ITCZ, but never in middle latitudes. On one occasion one aircraft encountered the same ozone transient twice: on its flight towards Bangkok and again when heading back to Paris several hours later. On another occasion, two different aircraft flying from South America towards Europe sampled the same ozone transient within a 30 minute interval and at a horizontal separation of 10 km. Measurement artefacts can therefore be ruled out as a possible cause for the high ozone mixing ratios.

On a spatial scale the observed events are small (5–80 km) and sharply marked. The proximity of the observed high-ozone values to equatorial deep convective systems, as indicated by aircraft deviations, turbulence, satellite images and humidity measurements (indicating the presence of ice particles), suggest that the observed phenomena are linked to these convective systems. We can exclude any local surface sources and orography-induced vertical transport as all selected events occur over the open ocean. In the free troposphere photochemically produced ozone, even in plumes from biomass burning, never exceeds levels of 100–150 p.p.b.v. (refs 6, 7). Cloud-charging processes caused by the dynamics in heavy cumulonimbus clouds can be a local source of ozone<sup>8</sup>: here ozone is produced by point discharges from water droplets in intense electrical fields<sup>9</sup> or by electrical discharging of a lightning flash<sup>10</sup>. In some field experiments in the vicinity of electrified clouds, high-ozone transients have been observed in the lower troposphere<sup>10,11</sup>. If the ozone transients observed during MOZAIC flights originate from discharging processes in cumulonimbus clouds, then the ozone is probably produced in the core of these clouds while the aircraft is flying outside the core in the divergence zone of the deep convective region. The ozone would then originate from air masses being detrained from the upward-moving convective flow. At present it is not clear if the dynamics within the observed deep



**Figure 1** Time series of: **a**, ozone mixing ratio; **b**, water vapour mixing ratio; **c**, temperature; **d**, barometric flight altitude variations; **e**, wind direction; **f**, wind speed. These time series were measured on MOZAIC flight Paris-São Paulo on 21 November 1994. The time of the encounter of ozone maxima are marked by vertical dashed lines on all graphs. This time series represents one hour of measurements during which the aircraft travelled 860 km from 7° 20' N, 27° 20' W to 0° 30' N, 30° 20' W. (data acquisition every 4 s, resulting in a spatial resolution of about 1 km). Before and after the event shown here, ozone mixing ratios were always well below 100 p.p.b.v. in tropical latitudes. Instrumentation was as follows: **a**, a dual-beam ultraviolet absorption instrument, Thermo-Electron model 49-103 (detection limit 2 p.p.b.v., overall precision  $\pm 2\%$ ); **b** and **c**, an AERODATA AD-F8-88 instrument installed in a Rosemount housing; **d**, **e**, **f**, Airbus A340 on-board instruments. Documentation of ten representative cases is available on the ftp site [ftp.aero.obs-mip.fr](ftp://ftp.aero.obs-mip.fr/pub/salsa/mozaic/SUPP1.PS) (file: /pub/salsa/mozaic/SUPP1.PS).

**Figure 2** Expanded-scale view of part of the ozone time series in Fig. 1a. Data points, represented by diamonds, are recorded every 4 s which corresponds to a horizontal resolution of  $\sim 1$  km. The location of the aircraft at the middle of this time series was approximately 3° 0' N, 28° 50' W.

convective clouds encountered concurrently with the observed ozone transients during the MOZAIC flights were sufficiently strong to induce cloud-charging processes with electric fields high enough to produce significant amounts of ozone.

An alternative and more likely explanation for our observations is that the high ozone mixing ratios have a stratospheric origin. Direct vertical downward transport could, for example, move ozone-rich air from the equatorial stratosphere into the upper troposphere. Recent measurements<sup>12,13</sup> show that ozone mixing ratios high enough to cause the ozone transients measured by the aircraft occur in mixed layers formed by tropical cumulus anvil outflow<sup>14</sup> in the stratosphere at about 18 km altitude. The observation of overshoots of cumulonimbus tops into the stratosphere<sup>15–17</sup> provides further evidence that the entrainment of stratospheric air into an anvil outflow is dynamically possible. Precipitation streaks loaded with ice particles falling from the anvil<sup>18–20</sup> can then transport ozone-rich air through the tropopause region into the upper troposphere. Further evaporative cooling and stronger downdrafts<sup>21</sup> may then move the air rapidly down to the aircraft's flight level at 10–12 km altitude. This mechanism requires continuous evaporative cooling to reduce the potential temperature of the air parcel from its originally high stratospheric value of  $\sim 400$  K to  $\sim 340$ – $350$  K at flight level. This can only be achieved by a continuous supply of ice particles growing in the anvil outflow, as proposed in Danielsen's stratospheric dehydration mechanism<sup>18,19</sup> with the largest particles falling close to the anvil outflow into the descending air parcel. The precise dynamical mechanism of that ozone downward transport still has to be established, and certainly needs more detailed sensing of the upper part of convective systems than available in the MOZAIC project. Another mechanism for vertical downward transport of stratospheric air into the troposphere has

been proposed by Newell *et al.*<sup>22</sup>. Although this process operates on much larger horizontal and temporal scales, it might play an initiating role in transferring stratospheric air several kilometres downwards before the actual convective event moves it down to flight level.

Instead of tracing their origin in the equatorial stratosphere, the observed ozone transients could also be the result of horizontal transport of ozone-rich air from the extratropical stratosphere followed by vertical convective transport to flight level. The central Atlantic, being downstream of the eastern major trough over the United States and at the tail end of the corresponding storm tracks, is a favoured region of extratropical wave-breaking far equatorward of the subtropical jet stream. Fragmentation of extratropical stratospheric-air intrusions<sup>23</sup> into elongated (2,000–3,000 km) and slender (200 km) streamers could be consistent with MOZAIC observations. This quasi-isentropic mechanism does not require cooling of the transported air mass (the 340–350 K isentropic layer in the middle latitudes is situated in the lower stratosphere where ozone mixing ratios are as high as those observed in the transients), but it is not yet clear whether thin streamers can extend into the equatorial region without undergoing significant mixing with tropospheric air. Meridional cross-sections of a climatologically averaged height of the 340–350 K isentropic layer indicate that the MOZAIC aircraft fly below this layer in the subtropics and enter it only in the equatorial latitude belt, in agreement with the observed locations of the ozone transients. Newell *et al.*<sup>24</sup> and Browell *et al.*<sup>25</sup> report ozone layers of stratospheric origin in the upper tropical troposphere, but with maximum ozone mixing ratios well below 100 p.p.b.v. The difficulty in determining whether this mechanism is at the origin of the observed ozone transients is that in such a long-range isentropic transport the air gradually loses its original



stratospheric characteristics by radiative processes (which act on the temperature and hence the potential vorticity). Although this air may keep its initial chemical composition, the loss of its stratosphere characteristics makes it difficult to distinguish from tropospheric air when using basic weather analysis charts.

The ozone transients encountered relatively often on the reported MOZAIC flights is a new finding that contrasts strongly with the near-zero ozone concentrations in about the same part of the upper troposphere observed by Kley *et al.*<sup>3</sup>, emphasizing the enormous variability of equatorial tropospheric ozone concentrations. We have put forward three possible explanations of the observed phenomena, without however being able to decide between them on the basis of current knowledge. Formation of ozone by cloud-charging processes at the observed scale (5–80 km) in the upper troposphere has so far not been observed, favouring a dynamical explanation. The possible existence of a mechanism of transport of stratospheric air into the upper equatorial troposphere is most intriguing and reflects our still poor knowledge of the dynamics of the upper part of the equatorial troposphere, particularly with regard to mass transfer between stratosphere and troposphere during deep convection and extratropical baroclinic wave-breaking extending into the equatorial belt. Furthermore, the existence of the ozone transients implies that the ozone budget of the equatorial troposphere is less well understood than has been assumed. □

Received 24 March; accepted 19 June 1997.

- Holton, J. R. *et al.* Stratosphere-troposphere exchange. *Rev. Geophys.* **33**, 403–439 (1995).
- Crutzen, P. J., Heidt, L. E., Krasnec, J. P., Polloc, W. H. & Seiler, W. Biomass burning as a source of atmospheric trace gases CO, H<sub>2</sub>, N<sub>2</sub>O, NO, CH<sub>3</sub>Cl and COS. *Nature* **282**, 253–256 (1979).
- Kley, D. *et al.* Observations of near-zero ozone concentrations over the convective Pacific: effects on air chemistry. *Science* **274**, 230–233 (1996).
- Marenco, A., Nedelec, P., Thouret, V. & Grouhel, C. in *DLR-Mitteilungen* (eds Schumann, U. & Wurzel, D.) **94-06**, 26–31 (DLR, Ober Pfaffen Hofen, 1994).
- Marenco, A. The MOZAIC programme. *EUROTRAC Newsl.* **17**, 2–7 (1996).
- Andreae, M. O. *et al.* Influence of plumes from biomass burning on atmospheric chemistry over the equatorial and tropical South Atlantic during CITE-3. *J. Geophys. Res.* **99**, 12793–12809 (1994).
- Delany, A. C., Haagensen, P., Walters, S., Wartburg, A. F. & Crutzen, P. J. Photochemically produced ozone in the emissions from large-scale tropical vegetation fires. *J. Geophys. Res.* **90**, 2425–2429 (1985).
- Griffing, G. W. Ozone and oxides of nitrogen production during thunderstorms. *J. Geophys. Res.* **82**, 943–950 (1977).
- Shlanta, A. & Moore, C. B. Ozone and point discharge measurements under thunderstorms. *J. Geophys. Res.* **77**, 4500–4510 (1972).
- Orville, R. E. Ozone production during thunderstorms, measured by the absorption of ultraviolet radiation from lightning. *J. Geophys. Res.* **72**, 3557–3561 (1967).
- Clarke, J. F. & Griffing, G. W. Aircraft observations of extreme ozone concentrations near thunderstorms. *Atmos. Environ.* **19**, 1175–1179 (1985).
- Vömel, H., Oltmans, S. J., Kley, D. & Crutzen, P. J. New evidence for the stratospheric dehydration mechanism in the equatorial Pacific. *Geophys. Res. Lett.* **22**, 3235–3238 (1995).
- Russel, P. B., Pfister, L. & Selkirk, H. B. The tropical experiment of the stratosphere-troposphere exchange project (STEP): science objectives, operations, and summary findings. *J. Geophys. Res.* **98**, 8563–8589 (1993).
- Lilly, D. K. Cirrus outflow dynamics. *J. Atmos. Sci.* **45**, 1594–1605 (1988).
- Roach, W. T. On the nature of the summit areas of severe storms in Oklahoma. *Q. J. R. Meteorol. Soc.* **93**, 318–336 (1967).
- Burnham, J. Atmospheric gusts—a review of the results of some recent research at the royal aircraft establishment. *Mon. Weath. Rev.* **98**, 723–734 (1970).
- Cornford, S. G. & Spavins, C. S. Some measurements of cumulonimbus tops in the pre-monsoon season in north-east India. *Meteorol. Mag.* **102**, 314–332 (1973).
- Danielsen, E. F. A dehydration mechanism for the stratosphere. *Geophys. Res. Lett.* **9**, 605–608 (1982).
- Danielsen, E. F. In situ evidence of rapid, vertical, irreversible transport of lower tropospheric air into the lower tropical stratosphere by convective cloud turrets and by large-scale upwelling in tropical cyclones. *J. Geophys. Res.* **98**, 8665–8681 (1993).
- Knollenberg, R. G., Kelly, K. & Wilson, J. C. Measurements of high number densities of ice crystals in the tops of tropical cumulonimbus. *J. Geophys. Res.* **98**, 8639–8664 (1993).
- Sun, J., Braun, S., Biggerstaff, M. I., Fovell, R. G. & Houze, R. A. Jr Warm upper-level downdrafts associated with a squall line. *Mon. Weath. Rev.* **121**, 2919–2927 (1993).
- Newell, R. E., Zhu, Y., Browell, E. V., Read, W. G. & Waters, J. W. Walker circulation and tropical upper tropospheric water vapor. *J. Geophys. Res.* **101**, 1961–1974 (1996).
- Appenzeller, C., Davies, H. C. & Norton, W. A. Fragmentation of stratospheric intrusions. *J. Geophys. Res.* **101**, 1435–1456 (1996).
- Newell, R. E. *et al.* Vertical fine-scale atmospheric structure measured from NASA DC-8 during PEM-West A. *J. Geophys. Res.* **101**, 1943–1960 (1996).
- Browell, E. V. *et al.* Large-scale air mass characteristics observed over Western Pacific during summertime. *J. Geophys. Res.* **101**, 1691–1712 (1996).

**Acknowledgements.** We thank Air France A340 pilot De Boysson for eye-witness information from MOZAIC flights in the tropics; we also acknowledge discussions with many of our colleagues. On behalf of the MOZAIC programme, we thank Air France, Lufthansa, Austrian Airlines and Sabena, who agreed to carry the MOZAIC equipment free of charge. This work was supported by the European Commission DG XII and Centre National de la Recherche Scientifique (CNRS-INSU); satellite images were made available by NOAA-SAA.

Correspondence should be addressed to K.S. (e-mail: suhk@aero.obs-mip.fr). Additional information on the MOZAIC programme is available on <http://www.cnrm.meteo.fr:8000/mozaic/>

## Conspecific sperm precedence in *Drosophila*

Catherine S. C. Price

Department of Ecology and Evolution, University of Chicago,  
1101 East 57th Street, Chicago, Illinois 60637, USA

Traits that influence the interactions between males and females can evolve very rapidly through sexual selection<sup>1</sup> or sexually antagonistic coevolution<sup>2</sup>. Rapid change in the fertilization systems of independent populations can give rise to reproductive incompatibilities between populations<sup>3,4</sup>, and may contribute to speciation<sup>5</sup>. Here I provide evidence for cryptic reproductive divergence among three sibling species of *Drosophila* that leads to a form of postmating isolation. When a female mates with both a conspecific and a heterospecific male, the conspecific sperm fertilize the vast majority of the eggs, regardless of the order of the matings. Heterospecific sperm fertilize fewer eggs after these double matings than after single matings. Experiments using spermless males show that the seminal fluid of the conspecific male is largely responsible for this conspecific sperm precedence. Moreover, when two males of the same species mate sequentially with a female from a different species, a highly variable pattern of sperm precedence replaces the second-male sperm precedence that is consistently found within species. These results indicate that females mediate sperm competition, and that second-male sperm precedence is not an automatic consequence of the mechanics of sperm storage.

Whenever a female mates with more than one male during a single fertile period, the opportunity for sperm competition is created<sup>6</sup>. A male's fitness on mating will be greatly influenced by his ability to interfere with resident sperm or to defend his sperm against interference by subsequent males. Male traits affecting sperm competition are therefore expected to be under strong selection. Female fitness is also influenced by multiple mating<sup>7</sup>, and females can be directly harmed by male adaptations for sperm competition<sup>8</sup>, so females should be under strong selection to mediate sperm competition. Such strong selection sets up the opportunity for rapid divergence among populations in traits that influence sperm competition. Nevertheless, the possibility of competition-dependent postmating, prezygotic reproductive isolation has not to my knowledge been investigated in organisms in which sperm competition has been extensively studied, such as *Drosophila*.

Conspecific sperm precedence (CSP) has been documented previously in grasshoppers<sup>9,10</sup>, crickets<sup>11</sup> and flour beetles<sup>12</sup>. It is not known, however, whether the reproductive isolation described in these species is strictly dependent upon sperm competition, as none of these studies directly compared the number of hybrids produced after a single observed heterospecific copulation with the number produced after one heterospecific and one conspecific copulation. Here I show that conspecific sperm preference among three species of *Drosophila* is a form of reproductive isolation that becomes apparent only after multiple matings, and is produced in large part by the seminal fluid of the conspecific male. Moreover, two independent experiments indicate that there is divergence among these species in traits that influence sperm competition.

The common outcome of sperm competition within most species of insects and birds is second-male precedence<sup>13</sup>. In *Drosophila*, females of many species mate with multiple males in nature, and in the laboratory the second male to mate typically fathers at least 85% of the offspring<sup>14</sup>. Although the mechanisms of sperm precedence in *Drosophila* remain unknown, the phenomenon appears to be influenced in large part by the seminal fluid of the second male<sup>15</sup>.

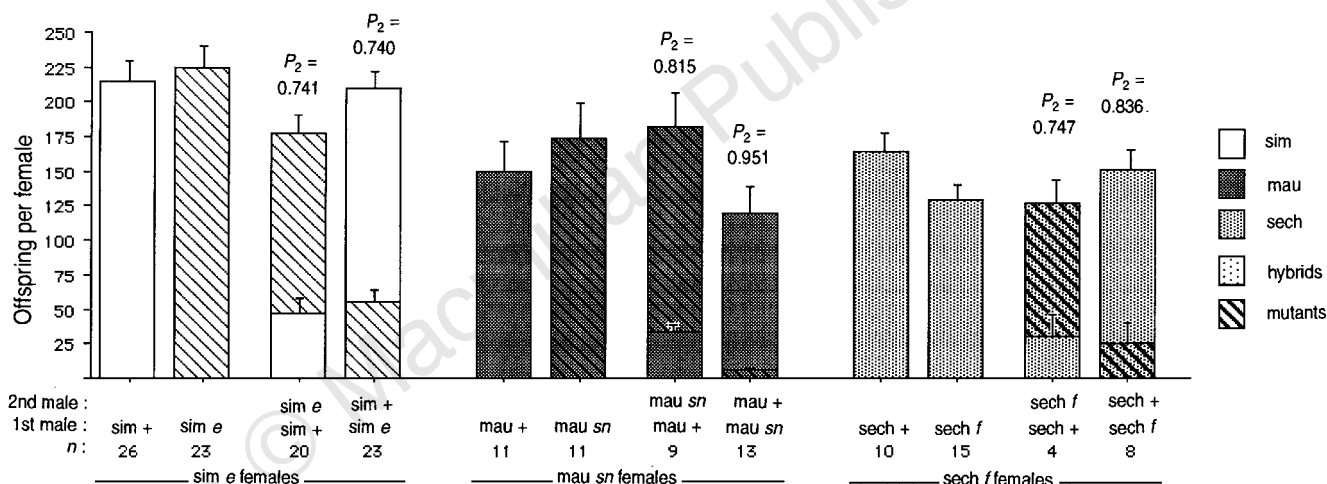
Furthermore, there is extensive genetic variation among males in sperm-competition ability<sup>16,17</sup>.

*Drosophila simulans* is a worldwide human commensal, whereas its sibling species *D. mauritiana* and *D. sechellia* are native only to Mauritius and the Seychelles, respectively, in the Indian Ocean. The three species are morphologically identical except for the shape of the male genital arch. *D. simulans* can be crossed reciprocally with either species in the laboratory, although *D. mauritiana* and *D. sechellia* females are both extremely reluctant to mate with *D. simulans* males. All interspecific matings produce sterile male and fertile female hybrids. My studies show that double matings within each of these three species produce the typical pattern of second-male sperm precedence (Fig. 1).

When a *D. mauritiana* male mates once with a *D. simulans* female and the copulation is long enough to result in sperm transfer, the mean number of hybrids produced ( $104 \pm 12.5$  (mean  $\pm$  s.e.)) is even greater than the mean number of offspring produced by a single conspecific *D. mauritiana* mating ( $80.8 \pm 14.9$ ). From these single matings alone we cannot determine whether heterospecific sperm are at any disadvantage in obtaining fertilizations. However,

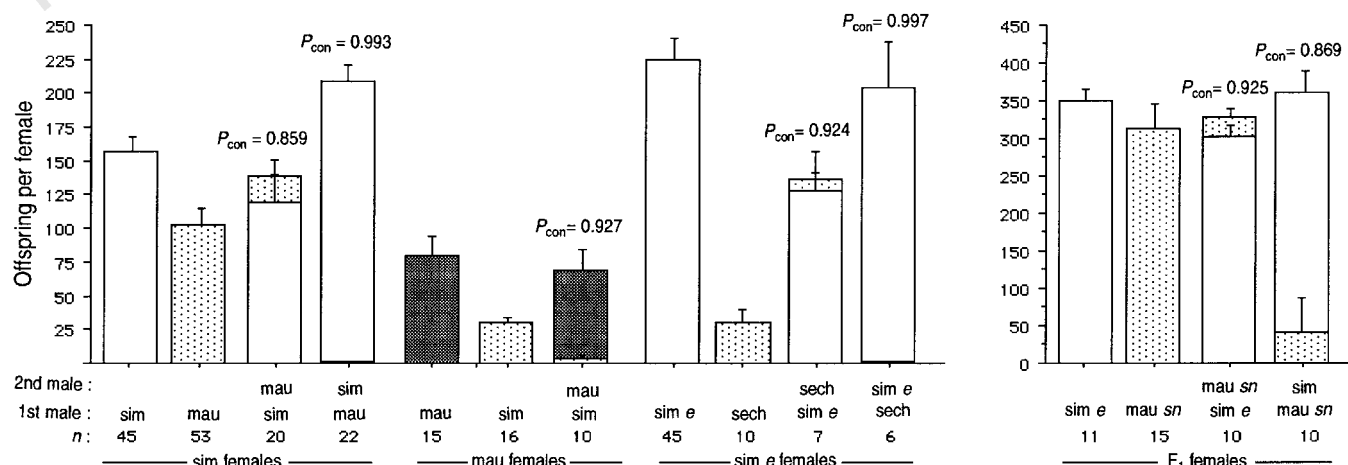
*D. mauritiana* males suffer a striking reproductive disadvantage when they compete with a *D. simulans* male for access to a *D. simulans* female's eggs (Fig. 2). When a *D. simulans* female mates first to a *D. simulans* male and then to a *D. mauritiana* male, the heterospecific male fathers on average only  $20 \pm 11.6$  individuals, compared to a mean of  $104 \pm 12.5$  hybrids produced after a single heterospecific mating (Mann-Whitney  $U$ ,  $P < 0.0001$ ). When the heterospecific male mates first, he suffers an even greater reproductive loss, fathering on average only  $1.1 \pm 0.6$  individuals (Mann-Whitney  $U$ ,  $P < 0.0001$ ). Conspecific sperm preference is quite strong, with the conspecific male fathering 86% of the offspring when he mates first, and 99% when he mates second. These results are not attributable to competition among larvae, as I detected no differences in viability between hybrid and *D. simulans* eggs placed at known proportions into food vials at densities similar to those resulting from the above matings (data not shown). These results were also confirmed using *D. mauritiana* and *D. simulans* from a second set of populations, indicating that conspecific sperm preference is not a strain-specific phenomenon (data not shown).

When a *D. mauritiana* female mates first to a *D. simulans* male



**Figure 1** Mean ( $\pm$  s.e.) number of offspring per female produced after single and double conspecific matings of *D. simulans*, *D. mauritiana* and *D. sechellia*. Offspring from the second mating are shown above those from the first mating.

The degree of second-male sperm precedence is given by  $P_2$ , the proportion of second-male offspring produced after remating. Abbreviations: sim, *D. simulans*; mau, *D. mauritiana*; sech, *D. sechellia*; e, ebony; sn, singed; f, forked.



**Figure 2** Mean ( $\pm$  s.e.) number of offspring per female produced after mating with a single conspecific male, a single heterospecific male, and both a conspecific and heterospecific male. The degree of conspecific sperm precedence is given

by  $P_{con}$ , the proportion of offspring sired by the conspecific male after remating.  $P_{sim}$  (given only for matings of  $F_1$  females) is the proportion of *D. simulans* offspring produced after remating.

and then to a *D. mauritiana* male, the outcome of competition is reversed, and the *D. mauritiana* (now conspecific) male fathers on average 93% of the offspring (Fig. 2). Because matings of the opposite order (with the heterospecific male second) were not possible owing to the extreme discriminatory behaviour of *D. mauritiana* females<sup>18</sup>, this result cannot be distinguished from simple second-male sperm precedence. Nevertheless, the results shown in Fig. 2 demonstrate that *D. mauritiana* sperm are not universally inferior competitors, but rather acquire a disadvantage within the reproductive tract of a heterospecific female. This competitive disadvantage indicates that *D. mauritiana* and *D. simulans* have diverged in both male and female traits that influence sperm competition.

The number of offspring produced by  $F_1$  hybrid females (*D. simulans*  $\times$  *D. mauritiana*) mated successively to males of the two parental species are shown in Fig. 2. Regardless of the order of the matings, the *D. simulans* male fathers the vast majority of the offspring. The female genes responsible for giving *D. simulans* sperm a competitive advantage therefore seem to be completely dominant.

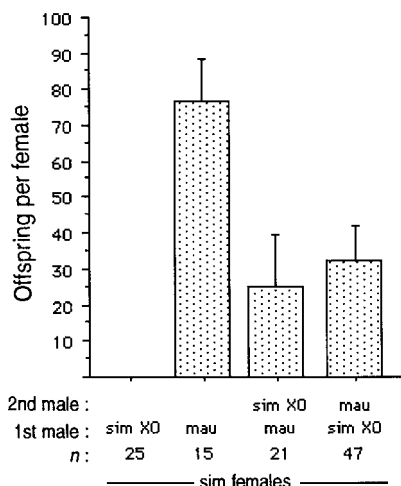
In another hybridization, single matings between *D. simulans* females and *D. sechellia* males produce on average only  $32.0 \pm 9.3$  hybrids, compared with a mean of  $256 \pm 16.8$  offspring produced after a single *D. simulans*  $\times$  *D. simulans* mating, and a mean of  $144 \pm 9.0$  produced after a single *D. sechellia*  $\times$  *D. sechellia* mating. The small number of hybrids is not caused by shortened copulation (data not shown), but may instead reflect problems with heterospecific sperm transfer, viability or storage. However, the production of hybrids is reduced even more severely when a *D. sechellia* male competes with a *D. simulans* male for access to *D. simulans* eggs (Fig. 2). When the *D. sechellia* male mates second, he fathers on average  $8 \pm 5.7$  offspring; and when he mates first he fathers on average only  $0.3 \pm 0.21$  offspring. Conspecific sperm preference between these species may in part be a consequence of low sperm viability, or other problems of *D. sechellia* sperm in the reproductive tract of a *D. simulans* female, but because the number of hybrids produced after double matings differs significantly from the number produced after single heterospecific matings (Mann–Whitney  $U$ ,  $P = 0.0194$ ), conspecific sperm preference in these species must once again be a competition-dependent phenomenon.

To determine the role of the conspecific seminal fluid in conspecific sperm preference, *D. simulans* females were mated to both a *D. mauritiana* male and a sterile XO (lacking a Y chromosome) *D.*

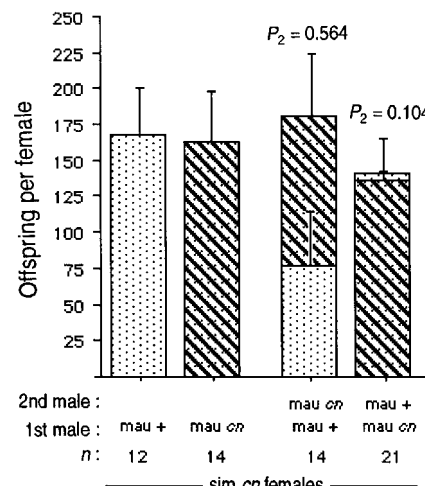
*simulans* male. These XO males have reduced testes and seminal vesicles, and do not transfer sperm; no sperm was observed anywhere in the reproductive tract of females dissected immediately after mating with an XO male ( $n = 7$ ). Regardless of the order of matings, the transfer of seminal fluid from the *D. simulans* male severely reduces the number of hybrid progeny produced (Fig. 3). This effect is not produced by the act of copulation itself, because double matings in which the XO copulation is interrupted shortly after initiation show no decrease in the number of hybrids produced (data not shown). Although the conspecific seminal fluid dramatically reduces the effectiveness of heterospecific sperm (a 61% reduction in offspring), this effect cannot entirely account for the reduction in hybrid offspring (95%) observed with conspecific sperm preference (Fig. 2).

When two *D. mauritiana* males are mated sequentially to a *D. simulans* female, the pattern of sperm precedence is strikingly different from both the conspecific sperm preference described above and the usual within-species second-male advantage (Fig. 4). The values of  $P_2$  (the proportion of offspring fathered by the second male after remating) range from 0 to 1, with 19 or 27 females showing first-male sperm precedence ( $P_2$  between 0 and 0.2), with two females showing sperm mixing ( $P_2$  between 0.5 and 0.6) and the remaining six females showing second-male sperm precedence ( $P_2$  above 0.85). The heterospecific female reproductive tract therefore changes the outcome of competition between two conspecific males. This result is consistent with the view that females strongly influence postmating, prefertilization events<sup>19</sup>, and indicates again that males and females have diverged between *D. simulans* and *D. mauritiana* for traits that influence sperm competition. This result also has implications for our understanding of the evolution of sperm precedence within species of *Drosophila*. Second-male sperm precedence is so nearly ubiquitous in the genus<sup>14</sup> that it might be considered a necessary consequence of the way in which multiple ejaculates are stored and used. The results described above, however, show that females can store sperm from two males of the same species in such a way as to produce the entire range of  $P_2$  values.

It has been suggested<sup>2</sup> that rapid antagonistic coevolution of male and female reproductive physiology may provide an “engine of speciation”. However, the predicted outcome of such an “engine”—postmating, prezygotic isolation—is often omitted from summaries of the types of reproductive isolation that can act between closely related species. Detecting competition-dependent postmating isolation, such as that described here, requires control of



**Figure 3** Mean (+ s.e.) number of offspring per female produced after mating *D. simulans* females with either a single *D. mauritiana* male, or both a *D. mauritiana* male and a sterile XO *D. simulans* male.



**Figure 4** Mean (+ s.e.) number of offspring per female produced after single and double matings of *D. mauritiana* males to *D. simulans* females; cn, cinnabar.

the number, order and duration of copulations, and determination of offspring paternity. Further studies of sperm competition involving taxa at different stages of divergence are needed to determine the frequency of conspecific sperm preference and the rapidity with which it arises. *Drosophila* is an ideal genus in which to pursue genetic and anatomic studies of conspecific sperm preference because there are mutant markers in many species, the fate of sperm can be followed easily within a female, and there are tools to produce males lacking sperm or various components of the seminal fluid<sup>20</sup>. □

## Methods

All stocks were reared at 24 °C with a 12-h light–dark cycle. Stock information is available from the author on request. Males and females were collected as virgins under CO<sub>2</sub> anaesthesia and stored in 8-dram food vials. Each female's first mating took place on the fourth day after eclosion, and her second mating (if any) occurred two days later. All males were four days old at the time of mating. Flies were transferred without anaesthesia into a food vial for observation, and 10–60 vials were observed simultaneously. All mating observations began within 1 h of lights coming on in the incubator, and lasted from 45 min to 10 h depending on the ease with which mating occurred. All copulations were observed and timed (unless otherwise noted), and males were removed from the observation vial within 5 min of copulation ending. Females failing to mate on the first day were discarded. Of the females that did mate successfully, a random subset was never given the opportunity to remate. Females that refused to remate were discarded. Females were stored individually in food vials for the two days between the first and second mating. They were transferred to a fresh vial for the second mating, and thereafter transferred to fresh vials every three days until they stopped laying fertile eggs. All offspring from each of these vials were reared to adulthood and scored for paternity. Paternity was determined by the presence of a mutant marker, except in the case of hybrids between *D. simulans* and *D. mauritiana*, which were distinguished by the shape of the male genital arch. F<sub>1</sub> females were obtained by crossing *D. simulans* ebony females with *D. mauritiana* singed males. XO males were produced by crossing virgin *D. simulans* females to males from a *D. simulans* attached-X, attached-XY stock. All male offspring from this cross lack only a Y chromosome, and so should produce normal seminal fluids<sup>21</sup>. All matings were performed as described above with the following exceptions. Copulations between *D. simulans* females and *D. mauritiana* males are often abnormally brief, resulting in interruptions of sperm transfer<sup>22</sup>. To control for this, all matings known not to result in sperm transfer were excluded from the study. Insemination was determined directly for single matings and first matings by the presence of larvae in the female's food vial. Insemination could not be assessed directly for second matings, but all matings shorter than those resulting in insemination after a single mating (<8 min) were excluded from the study. Matings of *D. mauritiana* females with *D. simulans* males could not be observed owing to the extreme discriminatory behaviour of the females. Instead, 10 virgin males and 10 virgin females were crowded together on the third day after eclosion, and left for 24 h at 24 °C. Females were transferred under CO<sub>2</sub> anaesthesia to individual vials on the morning of the fourth day after eclosion. Insemination was determined by the presence of larvae in the vials. Inseminated females were aged and re-mated to *D. mauritiana* males as described above.

Received 25 February; accepted 19 May 1997.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Rice, W. R. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**, 232–234 (1996).
3. Palumbi, S. R. & Metz, E. C. Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). *Mol. Biol. Evol.* **8**, 227–239 (1991).
4. Lee, Y. H. & Vacquier, V. D. The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol. Bull.* **182**, 97–104 (1992).
5. Rice, W. R. & Hostert, E. E. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47**, 1637–1653 (1993).
6. Parker, G. A. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* **45**, 525–567 (1970).
7. Pyle, D. W. & Gromko, M. H. Repeated mating by female *Drosophila melanogaster*: The adaptive importance. *Experientia* **34**, 449–450 (1978).
8. Chapman, T., Liddle, L. R., Kalb, M. J., Wolfner, M. F. & Partridge, L. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* **373**, 241–244 (1995).
9. Hewitt, G. M., Mason, P. & Nichols, R. A. Sperm precedence and homogamy across a hybrid zone in the alpine grasshopper *Podisma pedestris*. *Heredity* **62**, 343–353 (1989).
10. Bella, J. L., Butlin, R. K., Ferris, C. & Hewitt, G. M. Asymmetrical homogamy and unequal sex ratio

- from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies. *Heredity* **68**, 345–352 (1992).
11. Gregory, P. G. & Howard, D. J. A postinsemination barrier to fertilization isolates two closely related ground crickets. *Evolution* **48**, 705–710 (1997).
12. Wade, M. J., Patterson, H., Chang, N. W. & Johnson, N. A. Postcopulatory, prezygotic isolation in flour beetles. *Heredity* **72**, 163–167 (1994).
13. Smith, R. L. (ed.) *Sperm Competition and the Evolution of Animal Mating Systems* (Academic, New York, 1984).
14. Gromko, M. H., Gilbert, D. G. & Richmond, R. C. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 371–426 (Academic, New York, 1984).
15. Harshman, L. G. & Prout, T. Sperm displacement without sperm transfer in *Drosophila melanogaster*. *Evolution* **48**, 758–766 (1994).
16. Clark, A. G., Aguade, M., Prout, T., Harshman, L. G. & Langley, C. H. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139**, 189–201 (1995).
17. Hughes, K. A. Quantitative genetics of sperm competition in *Drosophila melanogaster*. *Genetics* **145**, 139–151 (1997).
18. David, J., Lemeunier, F., Tsacas, L. & Bocquet, C. Hybridation d'une nouvelle espèce, *Drosophila mauritiana* avec *D. melanogaster* et *D. simulans*. *Ann. Génét.* **17**, 235–241 (1974).
19. Eberhard, W. G. *Female Control: Sexual Selection by Cryptic Female Choice* (Princeton Univ. Press, New Jersey, 1996).
20. Herndon, L. A. & Wolfner, M. F. A *Drosophila* seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. *Proc. Natl Acad. Sci. USA* **92**, 10114–10118 (1995).
21. Ingman-Baker, J. & Candido, E. P. Proteins of the *Drosophila melanogaster* male reproductive system: two-dimensional gel patterns of proteins synthesized in the XO, XY, and XYY testis and paragonial gland and evidence that the Y chromosome does not code for structural sperm proteins. *Biochem. Genet.* **18**, 809–828 (1980).
22. Coyne, J. A. The genetics of an isolating mechanism between two sibling species of *Drosophila*. *Evolution* **47**, 778–788 (1993).

**Acknowledgements.** I thank K. C. Price, P. Rooney, K. Kyle, C. Kim and K. Dyer for technical assistance; J. Coyne for inspiration and help; and H. A. Orr, M. Turelli, M. Noor, N. Johnson and M. Wade for comments. This work was supported by an NSF predoctoral fellowship to the author, and by an NIH grant to J. Coyne.

Correspondence and requests for materials should be addressed to the author (e-mail: csprice@midway.uchicago.edu).

## Molecular evidence from retroposons that whales form a clade within even-toed ungulates

Mitsuru Shimamura\*, Hiroshi Yasue†, Kazuhiko Ohshima\*, Hideaki Abe\*, Hidehiro Kato‡, Toshiya Kishiro‡, Mutsuo Goto§, Isao Munechika|| & Norihiro Okada\*

\* Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Yokohama, Kanagawa 226, Japan

† Animal Genome Research Group, National Institute of Animal Industry, 2 Ikenodai, Kuisaki-machi, Inashiki-gun, Ibaraki 305, Japan

‡ Large Cetacean Section, National Research Institute of Far Seas Fisheries, 5-7-1 Orido, Shimizu, Shizuoka 424, Japan

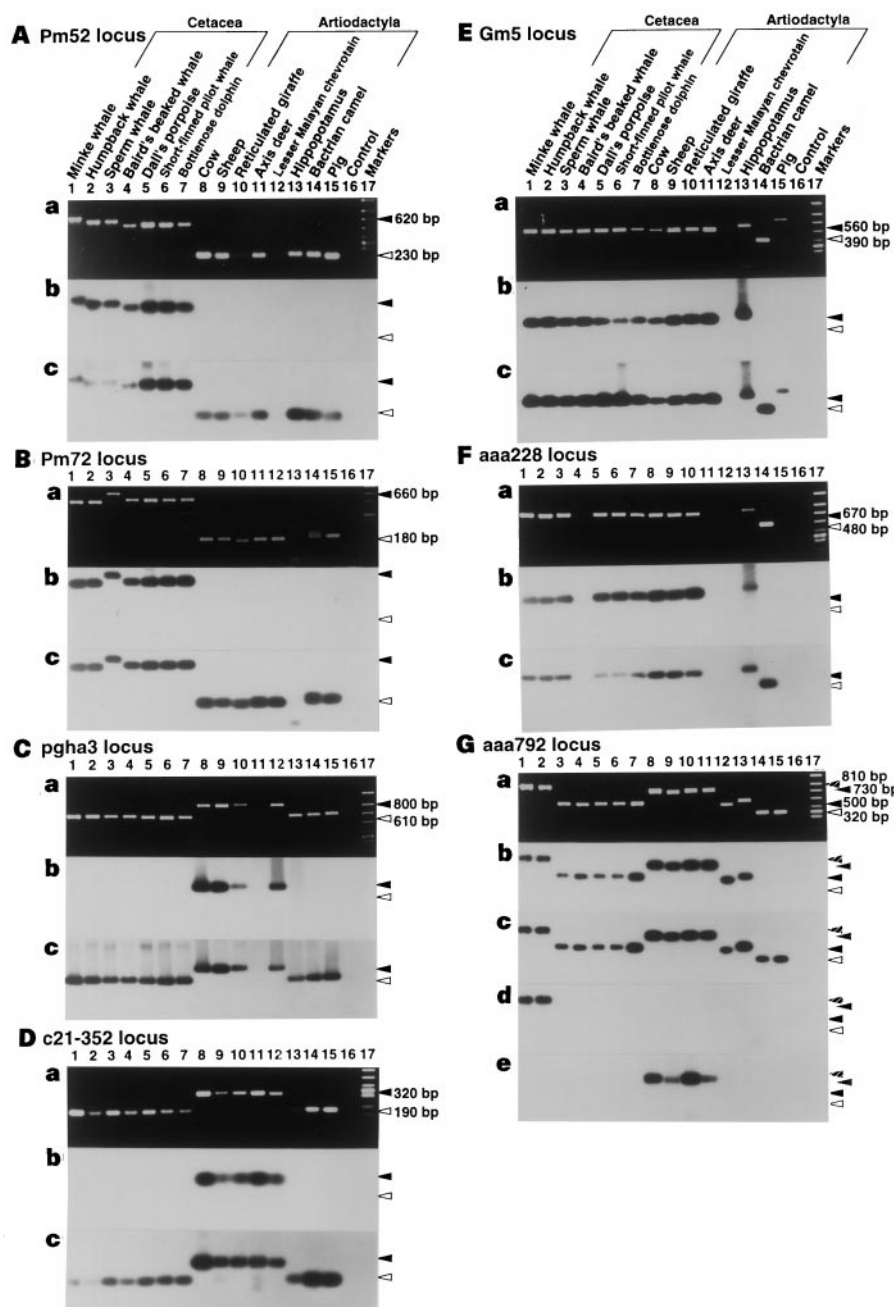
§ Genetic Ecology Section, The Institute of Cetacean Research, 4-18 Toyomi-cho, Chuo-ku, Tokyo 104, Japan

|| Chiba Zoological Park, 280 Minamoto-cho, Wakaba-ku, Chiba 264, Japan

The origin of whales and their transition from terrestrial life to a fully aquatic existence has been studied in depth. Palaeontological<sup>1,2</sup>, morphological<sup>3</sup> and molecular studies<sup>4–7</sup> suggest that the order Cetacea (whales, dolphins and porpoises) is more closely related to the order Artiodactyla (even-toed ungulates, including cows, camels and pigs) than to other ungulate orders. The traditional view that the order Artiodactyla is monophyletic has been challenged by molecular analyses of variations in mitochondrial and nuclear DNA<sup>5–7</sup>. We have characterized two families of short interspersed elements (SINEs) that were present exclusively in the genomes of whales, ruminants and hippopotamuses, but not in those of camels and pigs. We made an extensive survey of retropositional events that might have occurred during the divergence of whales and even-toed ungulates. We have characterized nine retropositional events of a SINE unit, each of which provides phylogenetic resolution of the relationships among whales, ruminants,







**Figure 2** Analysis of the seven loci at which a SINE unit(s) was inserted during the evolution of cetaceans, ruminants and hippopotamuses: **A**, Pm52; **B**, Pm72; **C**, pgba3; **D**, c21-352; **E**, Gm5; **F**, aaa228; **G**, aaa792. **a**, Products of PCR; **b**, **c**, results of hybridization experiments with different kinds of probe, namely, a unit sequence of the SINE (**b**) and the flanking sequence (**c**), respectively. In **G**, **d** and **e** show results of hybridization experiments with two different SINE probes, the CHR-2 SINE and the Bov-tA SINE, respectively.

cetaceans and even-toed ungulates. The first integration event, involving a CHR-1 SINE, occurred in a common ancestor of cetaceans, ruminants and hippopotamuses (Fig. 2G, a–c). The pattern of PCR products is shown in Fig. 2G, a. Hybridization experiments with the CHR-1 sequence (Fig. 2G, b) and the flanking sequence (Fig. 2G, c) as probe, respectively, showed that the SINE unit was integrated at the orthologous loci of the species designated above lanes 1–13, but not at those of the camel (lane 14) or the pig (lane 15). These results confirm the monophyly of cetaceans, ruminants and hippopotamuses, excluding camels and pigs. However, the lengths of fragments generated by PCR (Fig. 2G, a) varied among species (lanes 1–13), but we deduced from the sequences of the main fragments that the other two different kinds of SINE were involved in this locus. Experiments using new probes confirmed that a CHR-2 SINE was integrated in the lineage of the minke and humpback whales (Fig. 2G, d, lanes 1 and 2), and that a Bov-tA SINE was integrated in the lineage of the pecora (cows, sheep, deer and giraffes), indicating that the lineage forms a monophyletic

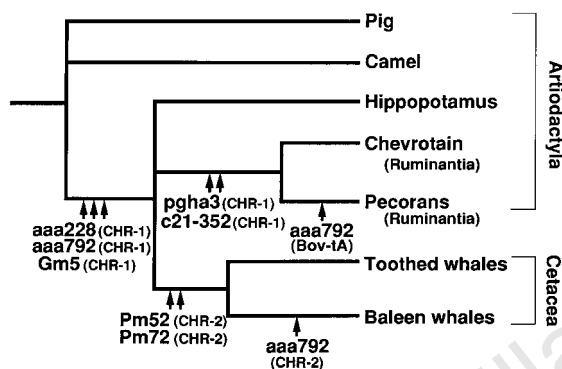
group (Fig. 2G, e, lanes 8–11). The sequences of major fragments are shown in Fig. 3.

All results for the seven loci are congruent (Fig. 4), and provide conclusive evidence for the paraphyly of the order Artiodactyla, which should include the order Cetacea, and for the paraphyly of the suborder Suiformes, from which hippopotamuses should be excluded. Hippopotamuses form a monophyletic group with cetaceans and ruminants.

The inclusion of cetaceans within the order Artiodactyla has been proposed previously<sup>5</sup>, as the Ruminantia/Cetacea clade with the Suiformes (pigs and peccaries) as an outgroup. The possibility of clustering the hippopotamus with the Cetacea has also been suggested<sup>6,7</sup>, even though hippopotamuses have traditionally been grouped with pigs and peccaries on morphological grounds<sup>16</sup>. However, careful reanalyses of available molecular data<sup>17–19</sup> indicate that the hypothesis of artiodactyl paraphyly was not supported convincingly from a statistical point of view. However, recent analyses of genes for milk casein<sup>7</sup> provided new, convincing support



**Figure 3** An alignment of sequences of the aaa792 locus in the cow (BT, *Bos taurus*), minke whale (BA, *Balaenoptera acutorostrata*), hippopotamus (HA, *Hippopotamus amphibius*) and bactrian camel (CB, *Camelus bactrianus*). Boxed sequences indicate direct repeats arising from duplication upon retroinsertion. The underlined sequences show the sequences used for primers. Bars indicate deletions. Nucleotides identical to those in the cow are indicated by dots.



**Figure 4** Phylogenetic relationships among cetaceans and artiodactyls, as deduced from the sites of insertion of SINEs. Arrows indicate the timing of insertion of SINEs. Types of SINE are shown in parentheses.

for the ((cetaceans, hippopotamuses), ruminants) tree, supporting a previous hypothesis<sup>5</sup>. Our analysis of SINE retrotranspositions seems to provide unambiguous support for that hypothesis.

The conclusions from our retropositional analysis are inconsistent with earlier morphologically based hypotheses<sup>16,20,21</sup>. Paleontological and morphological data suggest that modern whales originated from the Archaeocetes (primitive aquatic cetaceans), which first appeared in the early Eocene epoch<sup>22</sup>. The Archaeocetes are believed to have originated from mesonychians, which appeared before the Eocene<sup>20</sup>. However, the most primitive artiodactyls (Dichobunids) first appeared in the early Eocene, and the origin of nearly all the families of artiodactyls can only be traced back to the middle or the late Eocene<sup>23,24</sup>. Such a sequence of appearance of these animals is inconsistent with our molecular data. However, a recent calibration of molecular clocks suggests that divergences among orders of eutherian mammals can be traced back more than 100 Myr. Hence, diversification of avian and mammalian orders might not have been an adaptive radiation after the Cretaceous/Tertiary extinction event (65 Myr ago), but might have been correlated with the fragmentation of emergent land areas during the Cretaceous<sup>25</sup>. We believe that recent molecular data will lead to the reinterpretation by palaeontologists of many fossil records of Artiodactyla to match our conclusions. Extensive morphological reversals and convergences, as well as large gaps in the fossil record, will then have to be acknowledged. □

## Methods

**Polymerase chain reaction.** PCR was performed in a 50-μl reaction mixture containing 0.2 mM dNTP, 200 ng of primer, Tth buffer (final Mg<sup>2+</sup> concentration, 1.5 mM) and 1 unit of Tth DNA polymerase (Toyobo, Osaka). Annealing temperature was chosen from 49°C to 58°C. A portion of the PCR products was analysed by electrophoresis in an agarose gel containing 2% (w/v) Nusieve GTG and 1% (w/v) Seakem GTG (FMC BioProducts, Rockland, ME). Hybridization and washing were performed as described<sup>13</sup>.

**Sequences of primers for PCR.** Pm52 locus, 5' primer, 5'-TCCTGATTCC (C/T)CTGAACAAA-3', and 3' primer, 5'-GGG(G/A)AAGACT(C/T)CCA(G/A)

(C/T)TTTGAAT-3'; Pm72 locus, 5' primer, 5'-TTTAAAGCATGGCAGTTG-GATT(T/G/A)T-3', and 3' primer, 5'-GGATCTGTTTCTTACTTTGACC-3'; pgha3 locus, 5' primer, 5'-TCGGTGTGGTCTC(G/C)AC(C/T)CT-3', and 3' primer, 5'-TGC(C/T)CCAATCTATCA(G/A)TG(C/T)ATG-3'; c21-352 locus, 5' primer, 5'-GAGAATTCCTCTG(G/A)AT(G/A)GT(G/C)AC-3', and 3' primer, 5'-(G/A)(C/T)CCGCAGCTCCATGGA(G/A)CC-3'; Gm5 locus, 5' primer, 5'-GTAATGTGATTGGCTTAGTGC-3', and 3' primer, 5'-TCAGTCTCTGG-TGGCAGTCT-3'; aaa228 locus, 5' primer, 5'-GCTTGATACCTACCACTAT-GAA-3', and 3' primer, 5'-CCTGG(A/C)(A/T)GTCT(G/C)AATTTGCAC-3'; and aaa792 locus, 5' primer, 5'-TGTGGA(A/T)(G/T)NTG(G/C)CAGATT(T/A)AAAG-3', and 3' primer, 5'-CAGCCACTTGCTCTTCAATAGC-3'.

**Locus.** Of the seven loci described, Pm52 and Pm72 were newly isolated by cloning and sequencing from a genomic library of sperm whales (*Physeter macrocephalus*), and Gm5 was isolated from that of short-finned pilot whale (*Globicephala macrorhynchus*). The other four loci were found in bovine genomic sequences in the EMBL database, as follows (accession numbers, locus name, SINE family): (X00004, pgha3, CHR1); (M11267 and M13545, c21-352, CHR-1); (X64565 and S48112, aaa228, CHR-1); (X64565 and S48112; aaa792; CHR-1, CHR-2 and Bov-tA).

Received 31 January; accepted 28 May 1997.

- Gingerich, P. D., Smith, B. H. & Simons, E. L. Hind limbs of Eocene *Basilosaurus*: evidence of feet in whales. *Science* **249**, 154–157 (1990).
- Thewissen, J. G. M. & Hussain, S. T. Origin of underwater hearing in whales. *Nature* **361**, 444–445 (1993).
- Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125 (1992).
- Milinkovitch, M. C., Orti, G. & Meyer, A. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. *Nature* **361**, 346–348 (1993).
- Graur, D. & Higgins, D. G. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. *Mol. Biol. Evol.* **11**, 357–364 (1994).
- Irwin, D. M. & Arnason, U. Cytochrome b gene of marine mammals: Phylogeny and evolution. *J. Mamm. Evol.* **2**, 37–55 (1994).
- Gatesy, J., Hayashi, C., Cronin, M. A. & Arctander, P. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Mol. Biol. Evol.* **13**, 954–963 (1996).
- Weiner, A. M., Deininger, P. L. & Efstratiadis, A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**, 631–661 (1986).
- Schmid, C. & Marais, R. Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2**, 874–882 (1992).
- Okada, N. SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends Ecol. Evol.* **6**, 358–361 (1991).
- Okada, N. & Ohshima, K. In *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (ed. Marais, R. J.) 61–79 (Landes, Austin, TX, 1995).

12. Murata, S., Takasaki, N., Saitoh, M. & Okada, N. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl Acad. Sci. USA* **90**, 6995–6999 (1993).
13. Murata, S., Takasaki, N., Saitoh, M., Tachida, H. & Okada, N. Details of retropositional genome dynamics that provide a rationale for a generic division: The distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). *Genetics* **142**, 915–926 (1996).
14. Takahashi, H., Awata, T. & Yasue, H. Characterization of swine short interspersed repetitive sequences. *Anim. Genet.* **23**, 443–448 (1992).
15. Lenstra, J. A., van Bostel, J. A. F., Zwaagstra, K. A. & Schwerin, M. Short interspersed nuclear element (SINE) sequences of the Bovidae. *Anim. Genet.* **24**, 33–39 (1993).
16. Gentry, A. W. & Hooker, J. J. in *The Phylogeny and Classification of the Tetrapods* Vol. 2, *Mammals* (ed. Benton, M. J.) 235–272 (Clarendon, Oxford, 1988).
17. Adachi, J. & Hasegawa, M. Instability of quartet analyses of molecular sequence data by the maximum likelihood method: The Cetacea/Artiodactyla relationships. *Mol. Phylog. Evol.* **6**, 72–76 (1996).
18. Hasegawa, M. & Adachi, J. Phylogenetic position of cetaceans relative to artiodactyls: Reanalysis of mitochondrial and nuclear sequences. *Mol. Biol. Evol.* **13**, 710–717 (1996).
19. Philippe, H. & Douzery, E. The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. *J. Mamm. Evol.* **2**, 133–152 (1994).
20. Thewissen, J. G. M. Phylogenetic aspects of cetacean origins: a morphological perspective. *J. Mamm. Evol.* **2**, 157–184 (1994).
21. Prothero, D. R., Manning, E. M. & Fischer, M. in *The Phylogeny and Classification of the Tetrapods* Vol. 2, *Mammals* (ed. Benton, M. J.) 201–234 (Clarendon, Oxford, 1988).
22. Fordyce, R. E. & Barnes, L. G. The evolutionary history of whales and dolphins. *Annu. Rev. Earth Planet. Sci.* **22**, 419–455 (1994).
23. Rose, K. D. Skeleton of *Diacodexis*, oldest known artiodactyl. *Science* **216**, 621–623 (1982).
24. Goltz, D. J. Eocene Artiodactyla of southern California. *Nat. Hist. Mus. Los Angeles County, Sci. Bull.* **26**, 1–85 (1976).
25. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).

**Acknowledgements.** We thank the Zoological Society of San Diego's Center and Y. Mukai in the Meat Hygienic Inspection Office in Ueda, Nagano prefecture for providing a sample of DNA from the lesser Malayan chevrotain and samples of DNA from calf, pig and sheep, respectively. This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science, Sports and Culture of Japan.

Correspondence and requests for materials should be addressed to N.O. (e-mail: nokada@bio.titech.ac.jp).

## Formation of olfactory memories mediated by nitric oxide

K. M. Kendrick, R. Guevara-Guzman, J. Zorrilla, M. R. Hinton, K. D. Broad, M. Mimmack & S. Ohkura

Department of Neurobiology, The Babraham Institute, Babraham, Cambridge CB2 4AT, UK

Sheep learn to recognize the odours of their lambs within two hours of giving birth, and this learning involves synaptic changes within the olfactory bulb<sup>1,2</sup>. Specifically, mitral cells become increasingly responsive to the learned odour, which stimulates release of both glutamate and GABA ( $\gamma$ -aminobutyric acid) neurotransmitters from the reciprocal synapses between the excitatory mitral cells and inhibitory granule cells<sup>1</sup>. Nitric oxide (NO) has been implicated in synaptic plasticity in other regions of the brain as a result of its modulation of cyclic GMP levels<sup>3–7</sup>. Here we investigate the possible role of NO in olfactory learning. We find that the neuronal enzyme nitric oxide synthase (nNOS) is expressed in both mitral and granule cells, whereas the guanylyl cyclase subunits that are required for NO stimulation of cGMP formation<sup>8</sup> are expressed only in mitral cells. Immediately after birth, glutamate levels rise, inducing formation of NO and cGMP, which potentiate glutamate release at the mitral-to-granule cell synapses. Inhibition of nNOS or guanylyl cyclase activity prevents both the potentiation of glutamate release and formation of the olfactory memory. The effects of nNOS inhibition can be reversed by infusion of NO into the olfactory bulb. Once memory has formed, however, inhibition of nNOS or guanylyl cyclase activity cannot impair either its recall or the neurochemical release evoked by the learned lamb odour. Nitric oxide therefore seems to act as a retrograde and/or intracellular messenger, being released from both mitral and granule cells to potentiate glutamate release from mitral cells by modulating cGMP contentra-

tions. We propose that the resulting changes in the functional circuitry of the olfactory bulb underlie the formation of olfactory memories.

NO may mediate synaptic plasticity changes in brain regions such as the hippocampus in order to influence both long-term potentiation (LTP) and memory formation<sup>3–7</sup>, but this idea is controversial<sup>9,10</sup>. In rodents, the inhibitory GABAergic granule and periglomerular cells of the primary sensory processing region for airborne odours, the olfactory bulb, contain large amounts of nNOS<sup>11</sup>, which converts arginine to citrulline and NO<sup>3,4</sup>. We investigated whether this could also be the case in sheep, and whether the increased sensitivity of mitral-to-granule cell synapses that results from learning the odours of their lambs might involve potentiation of glutamate release by NO through modulation of cGMP.

We cloned and sequenced the ovine nNOS gene and found that it was 88% homologous with human<sup>12</sup>, rat<sup>13</sup> and mouse<sup>14</sup> sequences (EMBL accession number, X99042). *In situ* hybridization and immunocytochemistry indicated that mitral as well as periglomerular and granule cells contained nNOS messenger RNA and protein (Fig. 1a). The dendrodendritic connections between the granule and mitral cells in particular contained large amounts of nNOS immunoreactivity. Only mitral cells expressed the mRNAs for the  $\alpha_a$  and  $\beta_1$  subunits of soluble guanylyl cyclase (Fig. 1a), which are both required for NO to cause cGMP formation<sup>8</sup>. The granule and periglomerular cells expressed low levels of only the  $\beta_1$  subunit. Thus NO probably only stimulates cGMP formation in mitral cells.

In brain, NO is released by glutamate acting on both NMDA (N-methyl-D-aspartate) and AMPA ((2-aminomethyl)phenylacetic acid) receptors<sup>3,4,15</sup>. In the mouse accessory olfactory bulb, activation of both receptor types is required for the formation of pheromonal olfactory memory<sup>16</sup> and both are present on the main olfactory bulb mitral and granule cells<sup>17</sup>. We used *in vivo* microdialysis to determine whether glutamate acts on both types of receptors to evoke NO and cGMP release in the olfactory bulb. Local retrodialysis infusions of NMDA or AMPA dose-dependently increased glutamate, GABA, NO and cGMP, although NMDA was more potent. These actions were blocked by specific receptor antagonists (Fig. 1b) and the NOS inhibitor L-nitroarginine (L-NARG). A selective inhibitor of NO-induced guanylyl cyclase, 1H-[1,2,4]oxadiazolo[4,3-a]quinoxalin-1-one (ODQ)<sup>18</sup> also prevented stimulation of cGMP, but not NO release (Fig. 1b). Both L-NARG and ODQ significantly reduced agonist-induced increases in glutamate and GABA, showing that agonist effects are partly mediated by NO and cGMP. Infusions of an NO donor, S-nitrosoacetylpenicillamine (SNAP), dose-dependently increased cGMP, glutamate and GABA; these effects were blocked by ODQ (Fig. 1b). Thus NO can act as a retrograde or intracellular messenger in the olfactory bulb by stimulating glutamate release through modulation of cGMP.

The importance of NO release in the olfactory bulb for plasticity changes underlying olfactory memory was investigated in *post partum* animals given bilateral local infusions by microdialysis probes of drugs targeting the NO signalling pathway. In control animals receiving either no treatment or an inactive enantiomer of L-NARG (D-NARG), which did not inhibit NOS, it was found that glutamate, GABA, noradrenaline, NO and cGMP all increased during the first 30 min after birth (Figs 2 and 3). Behavioural tests showed that all these animals formed an olfactory memory that allowed selective recognition of lambs (Fig. 4a). However, all animals treated with the ionotropic glutamate receptor antagonist  $\gamma$ -D-glutamylglycine (DGG), and NOS inhibitor L-NARG, or the guanylyl cyclase inhibitor ODQ, accepted their own and strange lambs equally, indicating that these agents completely prevented olfactory memory formation (Fig. 4a). The quality of maternal care shown by ewes towards their lambs during the two hours immediately *post partum* was unaffected by the drugs, confirming that

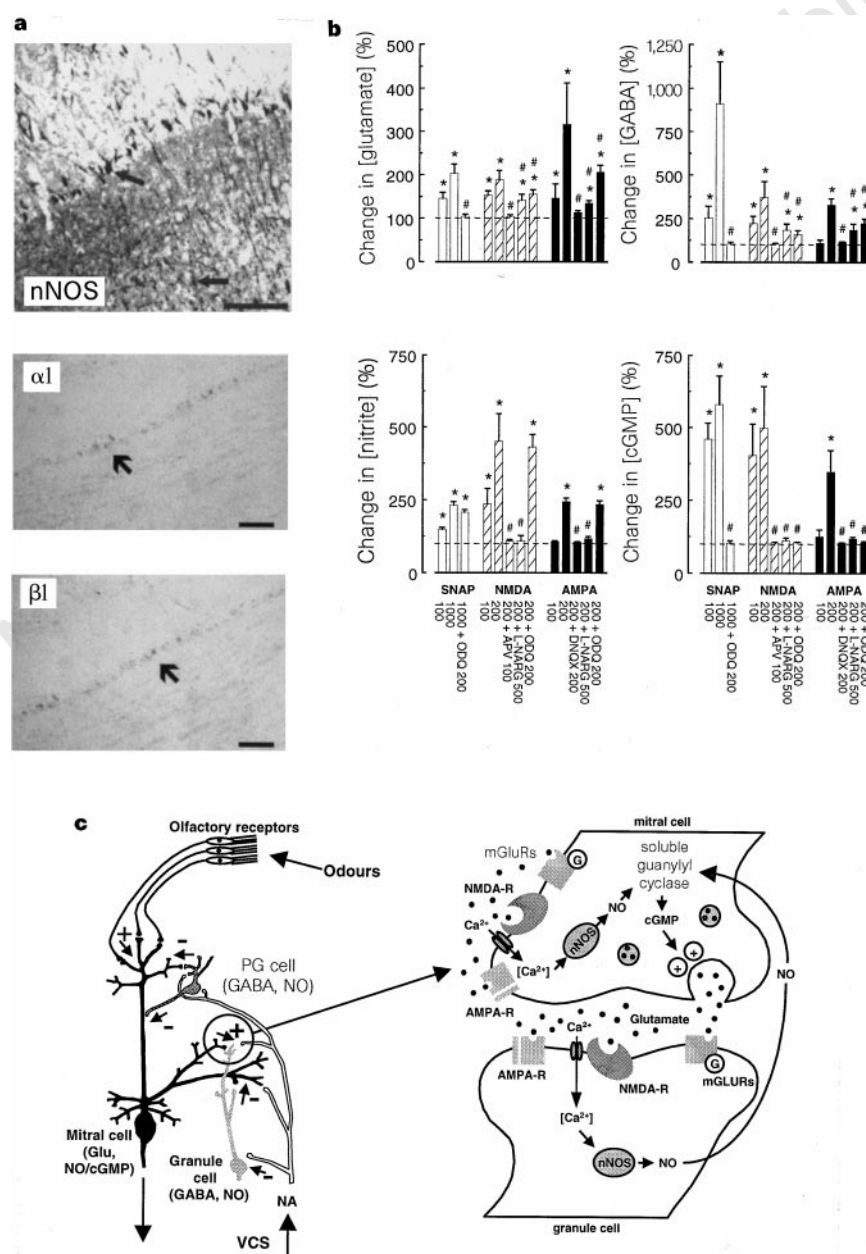


normal olfactory bulb function is not important for maternal behaviour *per se* in multiparous animals<sup>19</sup>. Both DGG and L-NARG prevented birth-induced increases in NO, cGMP, glutamate and GABA (Figs 2 and 3). With ODQ treatment, there was the normal increase in NO at birth, but there was no increase in cyclic GMP, glutamate or GABA (Fig. 2). Drugs targeting the NO signalling pathway therefore prevent both memory formation and the normal birth-induced potentiation of glutamate release. The birth-induced increase in noradrenaline release was unaffected by DGG, L-NARG and ODQ, even though memory formation was prevented. Thus noradrenergic action on olfactory memory formation<sup>20</sup> may be effected by the NO signalling pathway.

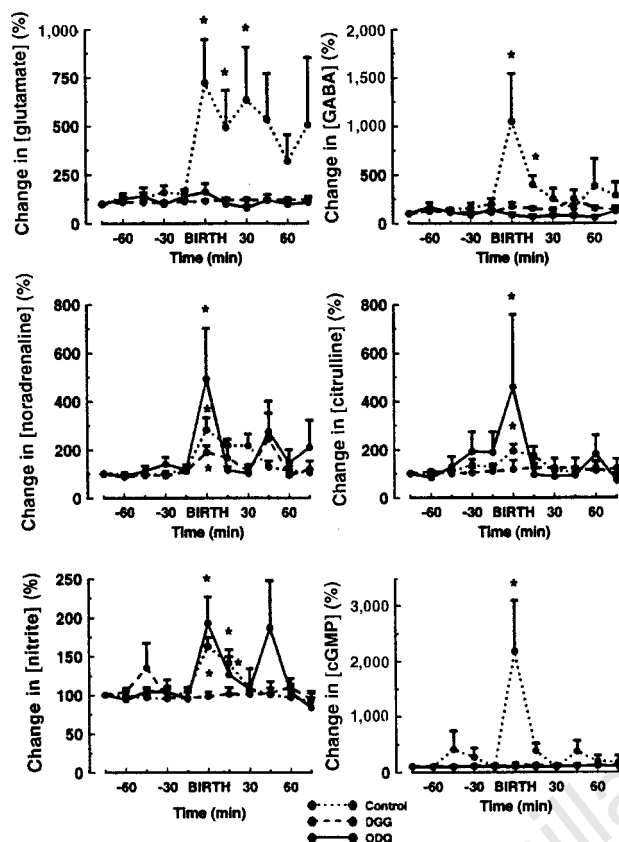
If NO does mediate plasticity changes in the olfactory bulb at birth, blockage of its release and olfactory memory formation by a NOS inhibitor might be reversed by local NO infusions. When a nitric oxide donor, SNAP, was co-infused with L-NARG for 1.5 h *post partum*, then memory formation was unaffected (Fig. 4a); however, the NO donor did not reverse ODQ-induced blockade of memory formation, indicating that NO released artificially must also act by stimulating cGMP accumulation in mitral cells. As infusing NO into the olfactory bulb for 1.5 h allowed a specific

olfactory memory to form, then the NO must promote plasticity changes only at those synapses between mitral and granule cells that are activated by both birth and specific lamb odours. If NO caused all synapses to be potentiated non-selectively, this would simply increase the overall level of noise in olfactory bulb circuits. Activity-dependent plasticity changes are also a characteristic of hippocampal synapses showing LTP<sup>21</sup> and the latter can be induced by NO infusion<sup>5</sup>.

None of the drugs that blocked olfactory memory formation did so irreversibly. After treatments ended at 8 h *post partum*, all animals learned to recognize their lambs selectively within 16 h. To test whether the NO signalling pathway is important for memory recall, we re-infused the drugs for a further 20 h. The ewes continued to recognize their own lambs' odours selectively (Fig. 4a). Thus this NO signalling pathway is not critical for memory recall and cannot be important for odour perception *per se*. Our finding that even DGG failed to block memory recall confirms that in the olfactory bulb, as in the hippocampus<sup>22</sup>, NMDA-receptor activation is required for memory formation but not for recall. It also indicates that the DGG infusions must have been restricted to the mitral and granule cell layers, because if they had spread into the glomerular



**Figure 1** **a**, Photomicrographs of nNOS protein expression in olfactory bulb (OB) mitral (top arrow in top panel) and granule (bottom arrow in top panel) cells, and emulsion autoradiography (bottom two panels) showing the mitral cell layer (arrow) expressing mRNA for the  $\alpha_1$  and  $\beta_1$  subunits of guanylyl cyclase. Scale bar, 200  $\mu$ m. **b**, Effects of SNAP, NMDA and AMPA on OB levels of glutamate, GABA, nitrite and cGMP alone, and after 6 h of treatment with selective NMDA, D,L-2-amino-5-phosphopentanoic acid (AP5) or AMPA/kainate, 6,7-dinitroquinoxaline-2,3-dione (DNQX) antagonists. Values are given as mean  $\pm$  s.e.m. percentage change, and drug concentrations are in  $\mu$ M. \* $P$  < 0.05 versus baseline (100%) and # $P$  < 0.05 versus the same dose within antagonist/inhibitor treatment (two-tailed Wilcoxon test). L-NARG treatment reduced basal nitrite levels by 40% and ODQ basal cGMP levels by 50%. **c**, Left, OB organization. Arrows show direction of inputs; and the plus and minus signs show excitatory and inhibitory influences. The mitral-to granule cell synapses circles are those where NO is proposed to act to potentiate glutamate release. Glu, glutamate; NA, noradrenaline; PG, periglomerular cell; VCS, vaginocervical stimulation. Right, proposed interactions between glutamate receptors and the NO/cyclic GMP pathway at the mitral-to-granule cell synapses during memory formation. AMPA-R, AMPA receptor; G, G protein; mGluRs, metabotropic receptors; NMDA-R, NMDA receptor.

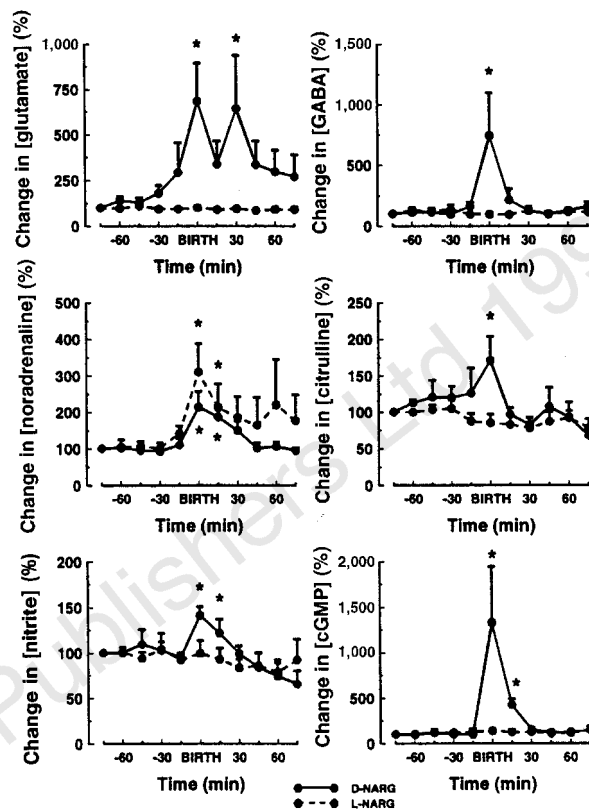


**Figure 2** Glutamate, GABA, noradrenaline, citrulline, nitrite and cGMP levels before, during, and after birth in control animals (dotted line), and in DGG (500  $\mu$ M; dashed line) and ODQ (200  $\mu$ M; solid line)-treated animals. \* $P < 0.05$ , compared with mean levels in the four samples before birth (two-tailed Wilcoxon test). Values are mean  $\pm$  s.e.m. percentage change relative to the time point 75 min before birth (set at 100%).

layer and blocked ionotropic receptors on the mitral cell apical dendrites, then the cells could not have responded to inputs from the olfactory nerve and odour perception would have been impaired.

After memory formation, own-lamb odours selectively increase glutamate and GABA<sup>1</sup> release, demonstrating that there is enhanced excitation of both the glutamatergic mitral cells and their associated inhibitory GABAergic granule-cell interneurons. If GABA is prevented from acting on GABA<sub>A</sub> receptors on mitral cells, sheep are unable to recognize their lambs selectively after memory formation<sup>2</sup>. If NO acts by promoting synaptic changes during memory formation that allow own-lamb odours to stimulate glutamate and GABA release more effectively, then drugs targeting this pathway should prevent potentiation of transmitter release, which is borne out by our results. Thus, whereas a 10-min exposure to own-lamb odours significantly increased glutamate and GABA in the olfactory bulb after memory formation in control animals, it did not do so in DGG, L-NARG or ODQ-treated animals in which memory formation was blocked (Fig. 4b). But when the NO donor SNAP prevented blockade of olfactory memory by L-NARG, then own-lamb odours did increase glutamate and GABA release (Fig. 4b). When memory formation occurred at the end of the drug treatment, own-lamb odours were able to evoke increased glutamate and GABA release, which was not prevented by their re-administration (Fig. 4b). Own-lamb odours never significantly increased NO or cGMP in these experiments, confirming their lack of direct involvement in memory recall.

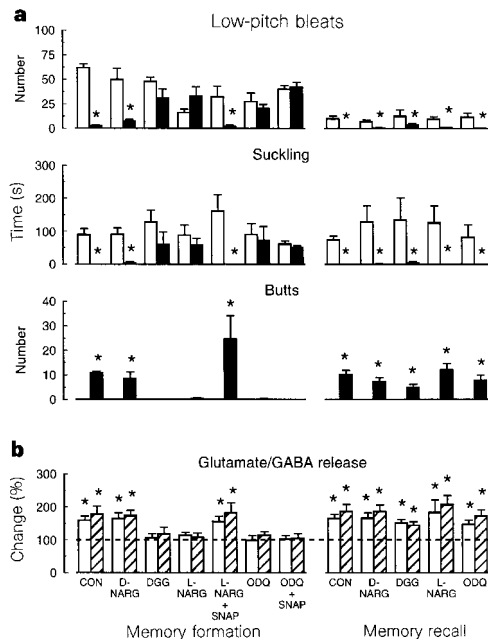
The fact that lamb-odour recognition after memory formation is



**Figure 3** Glutamate, GABA, noradrenaline, citrulline, nitrite and cGMP levels before, during, and after birth in control D-NARG (500  $\mu$ M; solid line) and L-NARG (500  $\mu$ M; dashed line)-treated animals. \* $P < 0.05$  compared with mean levels in the four samples before birth (two-tailed Wilcoxon test). Values are mean  $\pm$  s.e.m. percentage change relative to the time point 75 min before birth (set at 100%).

not associated with NO release or affected by the ionotropic receptor antagonist DGG suggests that the NMDA and AMPA receptors are not stimulated much. Odour-evoked increases in glutamate and GABA release may involve glutamate acting on metabotropic receptors (mGluRs) whose sensitivity has been increased by NO-induced plasticity changes during memory formation. In particular, mGluR1 is localized presynaptically on mitral cells as well as postsynaptically on granule cells<sup>23</sup>.

Our findings show that NO can mediate those plasticity changes in the brain that underlie memory formation through cGMP-dependent potentiation of glutamate release. We propose (Fig. 1c) that olfactory memory formation induced at birth involves several steps. First, vaginocervical stimulation during birth evokes noradrenaline release at the centrifugal synapses with olfactory bulb granule cells, which acts at  $\beta$ -noradrenergic receptors<sup>20</sup> to reduce GABA release at the granule-to-mitral cell synapses. This relieves the inhibition of the mitral cells, and glutamate released at their reciprocal synapses with granule cells acts on NMDA and/or AMPA receptors to stimulate NO release from both pre- and postsynaptic sites. Nitric oxide further potentiates glutamate release through modulation of cGMP, resulting in a long-term increase in the sensitivity of the mitral-to-granule cell synapses to glutamate. This positive-feedback cycle has parallels with hippocampal LTP<sup>5,21,22,24,25</sup>. When mitral cells responding to the lamb's odours are activated after memory formation, glutamate becomes more potent, acting by means of autoreceptors both to increase the excitation of these cells and subsequently to inhibit them further through GABA release potentiated from reciprocal granule-cell



**Figure 4** **a**, The effects of drug infusions in the olfactory bulb on acceptance (low-pitch bleating and suckling) and rejection behaviour (butting), as shown by maternal ewes towards their own (white bars) and strange (black bars) lambs. For memory formation, values are mean  $\pm$  s.e.m. of 4 tests (2, 4, 6 and 8 h *post partum*), and for recall, from 2 tests (44 and 46 h *post partum*). \* $P < 0.05$  versus own-lamb (two-tailed Wilcoxon test). DGG (50  $\mu$ M) and L-NARG (100  $\mu$ M) had no effects. **b**, Mean  $\pm$  s.e.m. percentage change in levels of glutamate (white bars) and GABA (hatched bars) in 15-min microdialysis samples taken from the olfactory bulb during 10-min exposures to own-lamb odours in the tests. \* $P < 0.05$  (two-tailed Wilcoxon).

synapses. This results in an enhanced pattern of phasic firing discharges from the cells in response to own-lamb odours. In this way, their output is more easily decoded by subsequent secondary and tertiary olfactory processing regions in the brain so that only own-lamb odours evoke maternal acceptance. □

## Methods

**Gene sequencing and histochemistry.** To identify the sheep nNOS gene, we generated a polymerase chain reaction (PCR) using specific primers to a conserved region of an exon-2 alignment between mouse, rat and human sequences. We sequenced 218 bases of the gene and used a digoxigenin-labelled riboprobe to show which olfactory bulb cells expressed it, by using *in situ* hybridization. We used 45-nucleotide antisense probes from  $\alpha 1$  (bases 1,197–1,240 and 1,377–1,423) and  $\beta 1$  (bases 1,168–1,213) bovine soluble guanylyl cyclase gene<sup>26</sup> subunits to localize their mRNA by autoradiography. The olfactory bulb cells and processes containing nNOS protein were revealed immunocytochemically using a polyclonal antibody (from Affiniti, Exeter; NC3000).

**In vivo microdialysis.** In the first experiment, 9 non-pregnant multiparous sheep were treated for 6 weeks with vaginal sponges containing oestradiol and progesterone to induce lactation and to allow maternal behaviour to be stimulated by vaginocervical stimulation. Microdialysis probes (4-mm membrane CMA-10 from CMA Microdialysis, Sweden) were inserted in the olfactory bulb by using surgically implanted guide tubes with Kreb's ringer solution (pH 7.4) pumping through them at 2  $\mu$ l min<sup>-1</sup>. Samples were collected at 15-min intervals. The effects of 15 min retrodialysis of NMDA, AMPA and SNAP from the microdialysis probes on NO (citrate, nitrite and nitrate<sup>15</sup>), cGMP and transmitter concentrations were measured alone or after 6 h retrodialysis infusion of L-NARG or ODQ, or 1-h infusions of AP5 or DNQX. Citrate, nitrite and nitrate and transmitters were measured by HPLC<sup>15</sup>, and cGMP was measured by ELISA (Cayman). Animals were tested at weekly

intervals for six weeks, with drug treatments given in a random order. For all experiments, concentrations were expressed as mean  $\pm$  s.e.m. percentage change from control (100%). Overall means  $\pm$  s.e.m. baseline levels measured were: cGMP, 122  $\pm$  26 pM; glutamate, 974  $\pm$  107 nM; GABA, 66  $\pm$  7 nM; nitrite, 669  $\pm$  33 nM; nitrate, 1,534  $\pm$  104 nM; noradrenaline, 421  $\pm$  70 pM. Bilateral microdialysis sampling was also carried out in the olfactory bulb of multiparous pregnant ewes as described<sup>1</sup>, starting 12–36 h *pre partum*. Neurochemical and behavioural assessments of olfactory memory formation and/or recall in control ( $n = 7$ ) or D-NARG ( $n = 6$ )-treated animals were compared with those receiving retrodialysis infusions of DGG (500  $\mu$ M,  $n = 6$ ; or 50  $\mu$ M,  $n = 4$ ), L-NARG (500  $\mu$ M,  $n = 6$ ; or 100  $\mu$ M,  $n = 4$ ) or ODQ (200  $\mu$ M,  $n = 5$ ). Drugs were given bilaterally for 6–24 h before birth and up to 8 h after it. Additional animals were given L-NARG ( $n = 4$ ) or ODQ ( $n = 3$ ) but SNAP was added for 1.5 h after birth. Drug treatments were repeated for 20 h starting 24 h *post partum*. Maternal responses to lambs were recorded for 2 h *post partum* (licking duration, low-pitched bleats, udder acceptance, suckling latency and duration). Behavioural assessments of olfactory memory formation were made at 2-h intervals starting 2 h after birth. First the ewe's own lamb was removed for 15 min, then a strange one was placed in her pen for 10 min, and maternal acceptance (low-pitch bleats, licking, acceptance of suckling, and suckling duration) and rejection (butting and refusing suckling attempts) behaviours scored. Her own lamb was returned 5 min later and behaviour scored for 10 min. Memory formation was considered to have failed if ewes showed full maternal acceptance (low-pitched bleating, licking and allowing suckling) of strange lambs and did not reject them by butting or refusing their suckling attempts on more than a single occasion; and if the frequency and duration of these behaviours towards strange lambs did not differ significantly from those shown towards their own lambs. Observers were  $>95\%$  in agreement on scores and treatments given blind.

**Tissue NOS levels.** After the experiments, animals were killed and their brains removed to confirm that the probes were correctly placed in the olfactory bulb and to quantify NOS activity by measuring the conversion of radiolabelled arginine to citrulline<sup>27</sup>. After L-NARG treatment (500  $\mu$ M), NOS activity was reduced by 62% ( $P < 0.01$ , *t*-test versus D-NARG controls). Levels in the hippocampus, amygdala, frontal cortex and cerebellum were unaffected. L-NARG (100  $\mu$ M) only reduced NOS activity in the olfactory bulb by 23% ( $P < 0.05$ ); DGG and ODQ had no effect in any region. A 10-min assay time is used in this standard method of assessing NOS activity<sup>27</sup> and may underestimate control NOS activity levels<sup>28</sup>.

Received 24 April; accepted 6 June 1997.

- Kendrick, K. M., Lévy, F. & Keverne, E. B. Changes in the sensory processing of olfactory signals induced by birth in sheep. *Science* **256**, 833–836 (1992).
- Kendrick, K. M. Neurobiological correlates of visual and olfactory recognition in sheep. *Behav. Proc.* **33**, 89–112 (1994).
- Garthwaite, J. & Boulton, C. L. Nitric oxide signalling in the central nervous system. *Annu. Rev. Physiol.* **57**, 683–706 (1995).
- Dawson, T. M. & Snyder, S. H. Gases as biological messengers: nitric oxide and carbon monoxide in the brain. *J. Neurosci.* **14**, 5147–5159 (1994).
- Arancio, O. *et al.* Nitric oxide acts directly in the pre-synaptic neuron to produce long-term potentiation in cultured hippocampal neurons. *Cell* **87**, 1025–1035 (1996).
- Zhuo, M., Small, S. A., Kandel, E. R. & Hawkins, R. D. Nitric oxide and carbon monoxide produce activity-dependent long-term synaptic enhancement in hippocampus. *Science* **260**, 1946–1950 (1993).
- Boulton, C. L., Southam, E. & Garthwaite, J. Nitric oxide-dependent long-term potentiation is blocked by a specific inhibitor of soluble guanylyl cyclase. *Neuroscience* **69**, 699–703 (1995).
- Furuya, T., Inagaki, S. & Takagi, H. Localisations of  $\alpha$ -1 and  $\beta$ -1 subunits of soluble guanylate cyclase in the rat brain. *Mol. Brain Res.* **20**, 335–344 (1993).
- Bannerman, D. M., Chapman, P. F., Kelly, P. A. T., Butcher, S. P. & Morris, R. G. M. Inhibition of nitric oxide synthase does not impair spatial learning. *J. Neurosci.* **14**, 7404–7414 (1994).
- Murphy, K. P. S. J., Williams, J. H., Bettache, N. & Bliss, T. V. P. Photolytic release of nitric oxide modulates NMDA receptor-mediated transmission but does not induce long-term potentiation at hippocampal synapses. *Neuropharmacology* **33**, 1375–1385 (1994).
- Kishimoto, J., Keverne, E. B., Hardwick, J. & Emson, P. C. Localization of nitric oxide synthase in the mouse olfactory and vomeronasal system: a histochemical, immunological and *in situ* hybridization study. *Eur. J. Neurosci.* **5**, 1684–1694 (1993).
- Fujisawa, H. *et al.* Expression of two types of nitric oxide synthase mRNA in human neuroblastoma cell lines. *J. Neurochem.* **63**, 140–145 (1994).
- Bredt, D. S. *et al.* Cloned and expressed nitric oxide synthase structurally resembles cytochrome P-450 reductase. *Nature* **351**, 714–718 (1991).
- Ogura, T., Yokoyama, T., Fujisawa, H., Kurashima, Y. & Esumi, H. Structural diversity of neuronal nitric oxide synthase mRNA in the nervous system. *Biochem. Biophys. Res. Commun.* **193**, 1014–1022 (1993).
- Kendrick, K. M. *et al.* NMDA and kainate-evoked release of nitric oxide and classical transmitters in the rat striatum: *in vivo* evidence that nitric oxide may play a neuroprotective role. *Eur. J. Neurosci.* **8**, 2619–2634 (1996).

16. Brennan, P. A. & Keverne, E. B. Impairment of olfactory memory by local infusions of non-selective excitatory amino acid receptor antagonists into the accessory olfactory bulb. *Neuroscience* **33**, 463–468 (1989).
17. Martin, L. J., Blackstone, C. D., Levy, A. I., Huganir, R. L. & Price, D. L. AMPA glutamate receptor subunits are differentially distributed in rat brain. *Neuroscience* **53**, 327–358 (1993).
18. Garthwaite, J. et al. Potent and selective inhibition of nitric oxide-sensitive guanylyl cyclase by 1H-[1,2,4]oxadiazolo[4,3-a]-1-one. *Mol. Pharmacol.* **48**, 184–188 (1995).
19. Lévy, F., Locatelli, A., Piketty, V., Tillet, Y. & Poindron, P. Involvement of the main but not the accessory olfactory system in maternal behavior of primiparous and multiparous ewes. *Physiol. Behav.* **57**, 97–104 (1995).
20. Lévy, F., Gervais, R., Kindermann, U., Orgeur, P. & Piketty, V. Importance of  $\beta$ -noradrenergic receptors in the olfactory bulb of sheep for recognition of lambs. *Behav. Neurosci.* **104**, 464–469 (1990).
21. Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
22. Morris, R. G. M. Synaptic plasticity and learning: selective impairment of learning and blockage of long-term potentiation *in vivo* by N-methyl-D-aspartate receptor antagonist, AP5. *J. Neurosci.* **9**, 3040–3057 (1989).
23. Van den Pol, A. N. Presynaptic metabotropic glutamate receptors in adult and developing neurons: autoexcitation in the olfactory bulb. *J. Comp. Neurol.* **359**, 253–271 (1995).
24. Doze, V. A., Cohen, G. A. & Maddison, D. V. Synaptic localization of adrenergic disinhibition in the rat hippocampus. *Neuron* **6**, 889–900 (1991).
25. Morris, R. G. M., Anderson, G. S., Lynch, M. & Baudry, M. Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist AP5. *Nature* **319**, 774–776 (1986).
26. Koesling, D. et al. The primary structure of the larger subunit of soluble guanylyl cyclase from bovine lung: homology between the two subunits of the enzyme. *FEBS Lett.* **266**, 128–132 (1990).
27. Brune, B. & Lapetina, E. G. Phosphorylation of nitric oxide synthase by protein kinase A. *Biochem. Biophys. Res. Commun.* **181**, 921–926 (1991).
28. Salter, M., Duffy, C. & Hazelwood, R. Determination of brain nitric oxide synthase inhibition *in vivo*: ex vivo assays of nitric oxide can give incorrect results. *Neuropharmacology* **34**, 327–334 (1995).

**Acknowledgements.** This work was supported in part by the Japanese Society for the Promotion of Science (S.O.) and by Consejo Nacional de Ciencia y Tecnología y Dirección General del Personal Académico (R.G.).

Correspondence and requests for materials should be addressed to K.M.K. (e-mail: keith.kendrick@bbsrc.ac.uk).

## Aggressiveness, hypoalgesia and high blood pressure in mice lacking the adenosine $A_{2a}$ receptor

Catherine Ledent\*, Jean-Marie Vaugeois†, Serge N. Schiffmann‡, Thierry Pedrazzini§, Malika El Yacoubi†, Jean-Jacques Vanderhaeghen‡, Jean Costentin†, John K. Heath||, Gilbert Vassart\*‡ & Marc Parmentier\*

\* IRIBHN, ‡ Laboratoire de Recherche sur les Neuropeptides, and † Service de Génétique Médicale, Université Libre de Bruxelles, Campus Erasme, 808 route de Lennik, B-1070 Bruxelles, Belgium

† IFRMP, Unité de Neuropsychopharmacologie Expérimentale, CNRS UPRESA 6036, Faculté de Médecine et de Pharmacie, Avenue de l'Université, 76803 Saint Etienne du Rouvray, France

§ Division of Hypertension, Lausanne University Medical School, CH-1011 Lausanne, Switzerland

|| CRC Growth Factors, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Adenosine is released from metabolically active cells by facilitated diffusion, and is generated extracellularly by degradation of released ATP. It is a potent biological mediator that modulates the activity of numerous cell types, including various neuronal populations, platelets, neutrophils and mast cells, and smooth muscle cells in bronchi and vasculature. Most of these effects help to protect cells and tissues during stress conditions such as ischaemia. Adenosine mediates its effects through four receptor subtypes: the  $A_1$ ,  $A_{2a}$ ,  $A_{2b}$  and  $A_3$  receptors<sup>1</sup>. The  $A_{2a}$  receptor ( $A_{2a}R$ )<sup>2,3</sup> is abundant in basal ganglia, vasculature and platelets, and stimulates adenylyl cyclase. It is a major target of caffeine, the most widely used psychoactive drug<sup>4</sup>. Here we investigate the role of the  $A_{2a}$  receptor by disrupting the gene in mice. We found that  $A_{2a}R$ -knockout ( $A_{2a}R^{-/-}$ ) mice were viable and bred normally.

Their exploratory activity was reduced, whereas caffeine, which normally stimulates exploratory behaviour, became a depressant of exploratory activity. Knockout animals scored higher in anxiety tests, and male mice were much more aggressive towards intruders. The response of  $A_{2a}R^{-/-}$  mice to acute pain stimuli was slower. Blood pressure and heart rate were increased, as well as platelet aggregation. The specific  $A_{2a}$  agonist CGS 21680 lost its biological activity in all systems tested.

The mouse  $A_{2a}R$  gene was cloned (Fig. 1). The coding region is interrupted by a single intron at a position corresponding to the second intracellular loop of the receptor. The deduced amino-acid sequence is 410 amino acids long, and shares 94.6 and 82.9% identity with the rat<sup>5</sup> and human<sup>6</sup> receptors, respectively. The  $A_{2a}R$  gene was inactivated in R1 embryonic stem (ES) cells by replacement of part of exon 1 with a neomycin-resistance ( $neo^r$ ) cassette (Fig. 1). Three neomycin-resistant clones were identified as bearing a recombined allele of the  $A_{2a}R$  gene. The correct integration of the mutant allele was demonstrated by Southern blotting, using various  $A_{2a}R$  and  $neo^r$  gene fragments as probes (not shown), and one of the clones was used to generate chimaeric mice. Two chimaeras bred with CD1 mice transmitted the mutant allele, and heterozygotes were bred for four generations on a CD1 background. Homozygous  $A_{2a}R^{-/-}$  mice were obtained with the expected mendelian frequency, indicating the lack of *in utero* or postnatal lethality. The absence of functional  $A_{2a}R$  transcripts was confirmed by northern blotting (not shown) and *in situ* hybridization on brain tissue (Fig. 2a), using probes included in the deleted sequence of exon 1. The lack of functional  $A_{2a}R$  was confirmed by binding assays using the specific  $A_{2a}$ -agonist [<sup>3</sup>H]CGS 21680 on brain slices (Fig. 2b) and membrane preparations from striatum (not shown).

$A_{2a}R^{-/-}$  mice appeared to be normal and bred normally. The body mass was significantly higher for knockout mice of both sexes from 3 months of age ( $39.8 \pm 0.8$  g and  $33.9 \pm 0.5$  g for 7-month-old  $A_{2a}R^{-/-}$  and  $A_{2a}R^{+/+}$  female mice respectively; mean  $\pm$  s.e.m.,  $n = 50$ ,  $P < 0.0001$ ), an unexplained observation. Routine histology on brain and other organs did not detect any abnormalities or loss of cells. Particular care was taken to investigate striatum, the brain area expressing the highest level of  $A_{2a}$  receptor. The  $A_{2a}R$  gene is expressed in basal ganglia from gestational day 14 in rat, and transient expression occurs in other developing brain structures<sup>7</sup>. *In situ* hybridization using a probe derived from exon 2 (detecting an abnormal transcript that included the  $neo^r$  cassette in  $A_{2a}R^{-/-}$  mice) labelled the striatum and olfactory tubercles similarly in  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice (Fig. 2a), suggesting that striatal neurons normally expressing the  $A_{2a}$  receptor develop normally and persist in adult mutant mice. Despite the proposed role of adenosine as a trophic factor for striatal and other neurons<sup>8</sup>, the  $A_{2a}$  receptor does not appear to be essential for neuron differentiation, migration or survival. However, undetected adaptative changes during development could contribute to the functional alterations described below. In the striatum,  $A_{2a}R$  is coexpressed with postsynaptic dopamine  $D_2$  receptors in GABAergic striatopallidal neurons, and regulates the expression of the proenkephalin gene.  $A_{2a}R$  is not present in striatonigral neurons that express the protachykinin and dopamine  $D_1$  receptor genes<sup>9</sup>. Using *in situ* hybridization, we found that protachykinin transcripts were decreased by  $30.1 \pm 2.1\%$  in  $A_{2a}R^{-/-}$  mice (mean  $\pm$  s.e.m.;  $n = 15$ ,  $P < 0.0001$ ), whereas proenkephalin transcripts were decreased by  $12.3 \pm 2.5\%$  ( $P < 0.002$ ) (Fig. 2a). The decrease in proenkephalin transcripts could result directly from the gene defect<sup>9,10</sup>. Decreased protachykinin gene expression suggests adaptative cortical changes that could involve a reduction in the activity of glutamatergic corticostriatal projections<sup>11</sup>.

Given the role of adenosine in modulating central functions through the  $A_{2a}$  receptor in basal ganglia and other brain areas<sup>1</sup>, knockout and wild-type mice were subjected to behavioural tests investigating exploratory activity, anxiety, aggressiveness and noci-

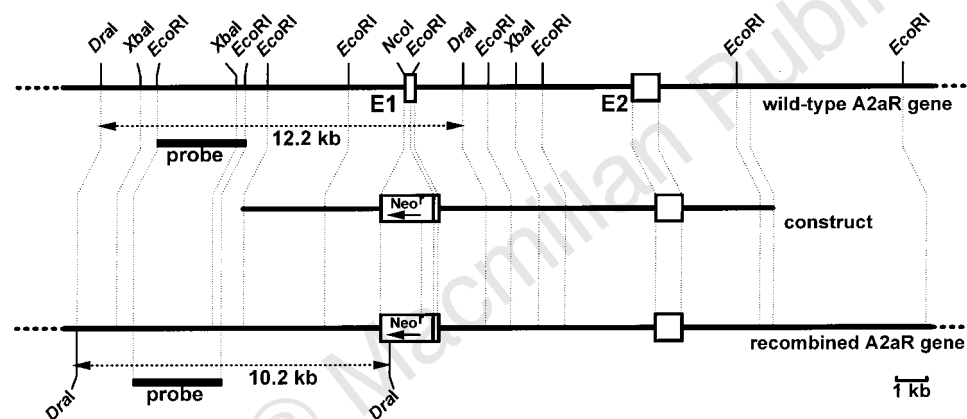


ceptive thresholds. Exploratory behaviour (horizontal and vertical components assayed in an open field) was reduced in  $A_{2a}R^{-/-}$  mice (Fig. 3a). Increased anxiety of mice (see below) when placed in a novel environment could contribute to the decreased exploratory behaviour. The persistence of decreased activity during the whole observation periods suggests, however, that it is not the only factor. The progressive decrease of activity over time was not affected. As expected, treatment with the  $A_{2a}$  agonist CGS 21680 strongly reduced locomotor activity in  $A_{2a}R^{+/+}$  mice, whereas the non-selective adenosine-receptor antagonist caffeine had a mild stimulatory effect (Fig. 3b). Also as expected, CGS 21680 had no significant effect on  $A_{2a}R^{-/-}$  mice. Caffeine had a strong depressant effect on knockout mice, probably resulting from the other biological effects of caffeine, such as blockade of  $A_1$  receptors<sup>4</sup>. The loss of locomotor stimulatory effects of caffeine in knockout mice demonstrates that the blockade of  $A_{2a}R$  is indeed responsible for its action in control mice. The decreased activity of  $A_{2a}R^{-/-}$  mice does not mimic the acute but rather the chronic blockade of the receptor by caffeine<sup>12</sup>. Adaptive processes modulating the activity of corticostriatal projections may be involved in these situations.

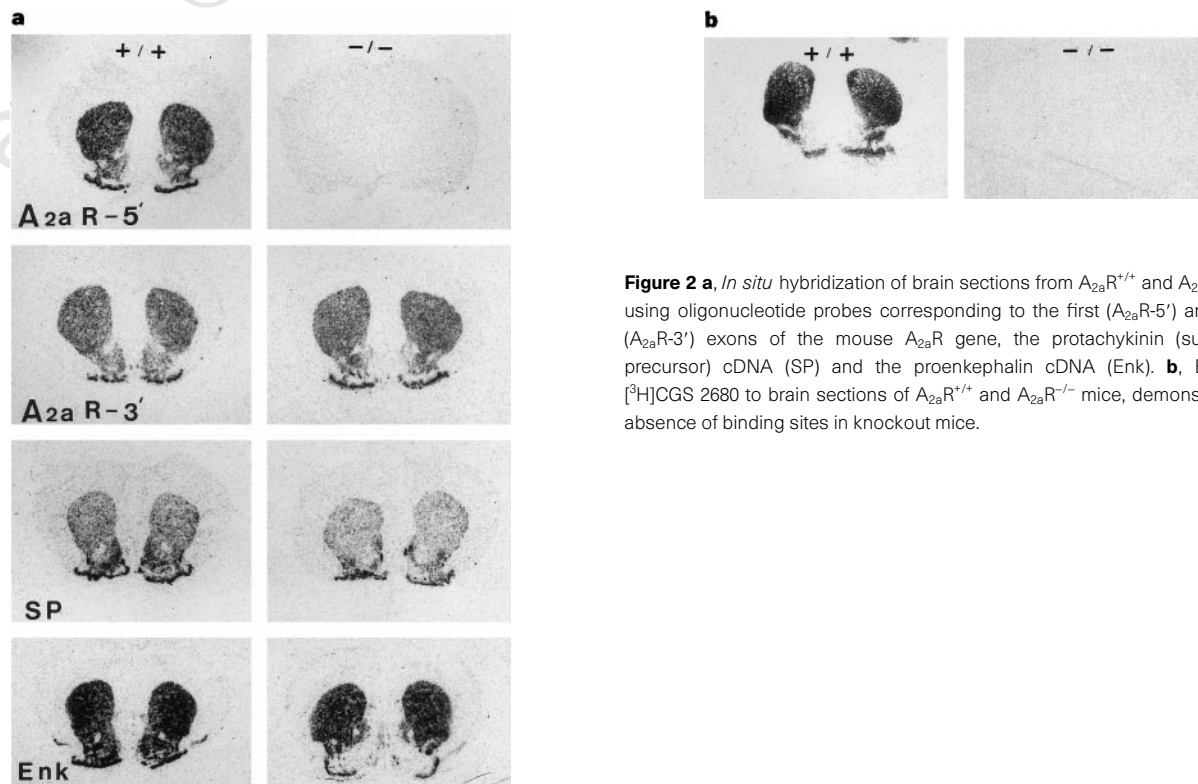
$A_{2a}R^{-/-}$  mice appeared to be more anxious than  $A_{2a}R^{+/+}$  mice (Fig.

3c): they spent more time in the closed arms of the plus-maze and in the black compartment of the two-compartment test. Caffeine increases anxiety<sup>13</sup> and adenosine-receptor stimulation has been correlated with anxiolytic-like behaviour, although  $A_1$  receptors have been considered more important in mediating this action<sup>14</sup>.  $A_{2a}R^{-/-}$  mice also appeared more aggressive. Aggressive behaviour was first suspected from the fact that bites were much more frequent among male  $A_{2a}R^{-/-}$  mice kept in the same cage than among controls. Increased aggressiveness was confirmed in the resident-intruder test<sup>15</sup>. Resident  $A_{2a}R^{-/-}$  males were much more aggressive towards intruders than  $A_{2a}R^{+/+}$  controls (Fig. 4), whereas there was no difference among females (not shown). This finding agrees with a report<sup>16</sup> showing that adenosine analogues inhibit fighting in male mice.

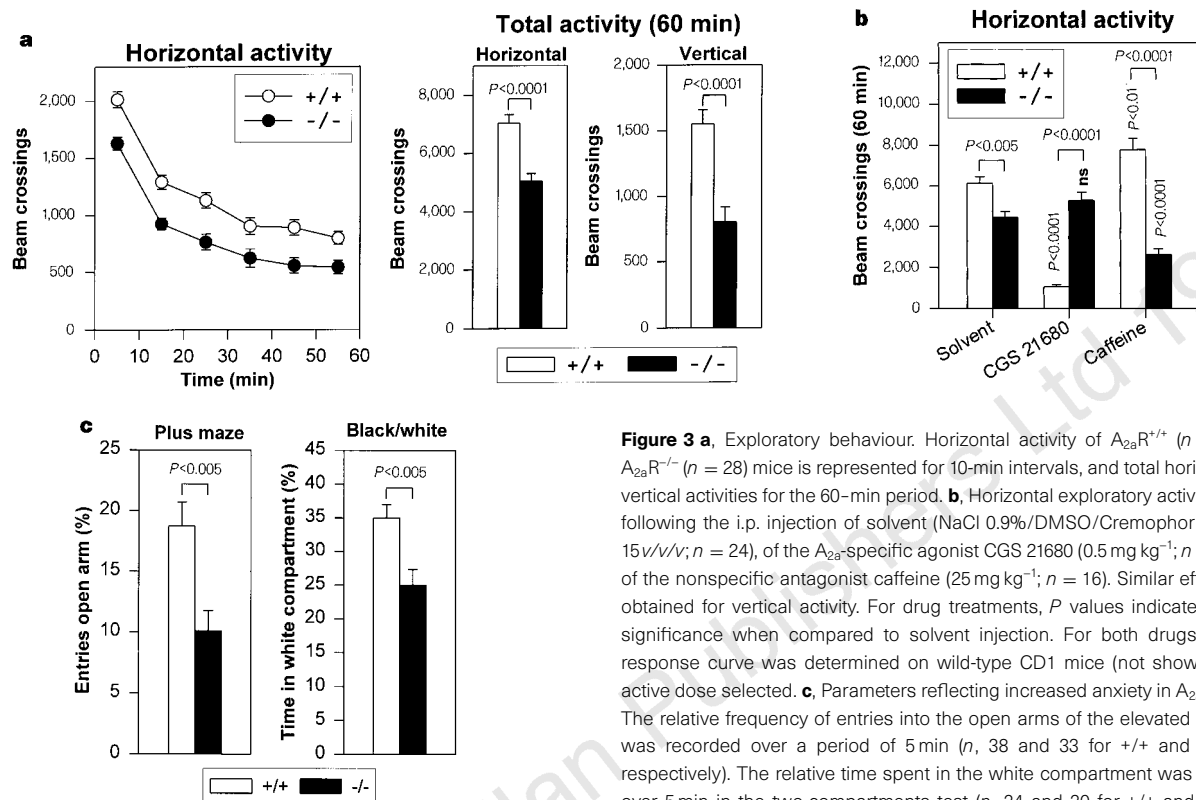
Adenosine is a modulator of nociceptive pathways, both at the peripheral and spinal levels<sup>17</sup>. Adenosine is released from spinal sites following opioid stimulation<sup>18</sup> and both  $A_1$  and  $A_{2a}$  receptors are involved in the antinociceptive effect of spinal adenosine<sup>19</sup>. At the periphery, direct stimulation of nociceptive nerve terminals by adenosine through  $A_{2a}$  receptors has been reported<sup>20</sup>, but adenosine also has complex effects (both stimulatory and inhibitory) on the



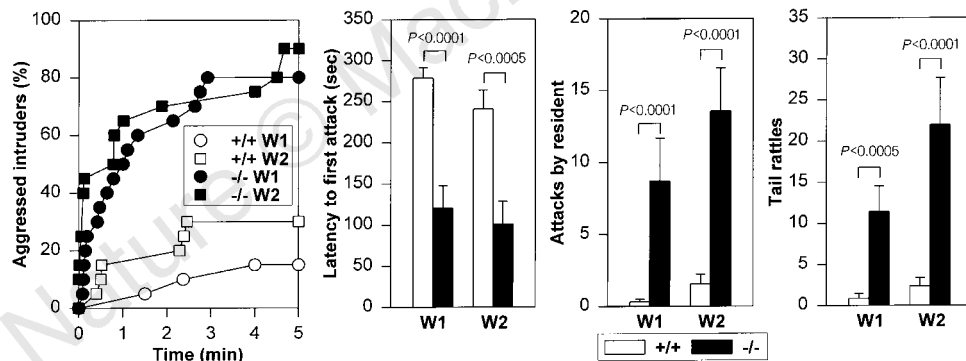
**Figure 1** Organization of the mouse  $A_{2a}R$  gene, the targeting construct, and the allele resulting from homologous recombination. Coding sequences of first (E1) and second (E2) exons are represented by boxes. Not all *DraI* and *NcoI* sites are known. The 2.8-kb *EcoRI* fragment used as a probe for the screening of recombinant alleles is shown as well as the size of the detected fragments following *DraI* digestion.



**Figure 2 a**, *In situ* hybridization of brain sections from  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice, using oligonucleotide probes corresponding to the first ( $A_{2a}R-5'$ ) and second ( $A_{2a}R-3'$ ) exons of the mouse  $A_{2a}R$  gene, the protachykinin (substance-P precursor) cDNA (SP) and the proenkephalin cDNA (Enk). **b**, Binding of [ $^3$ H]CGS 2680 to brain sections of  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice, demonstrating the absence of binding sites in knockout mice.



**Figure 3** **a**, Exploratory behaviour. Horizontal activity of  $A_{2a}R^{+/+}$  ( $n = 32$ ) and  $A_{2a}R^{-/-}$  ( $n = 28$ ) mice is represented for 10-min intervals, and total horizontal and vertical activities for the 60-min period. **b**, Horizontal exploratory activity (60 min) following the i.p. injection of solvent (NaCl 0.9%/DMSO/Cremophor EI, 70/15/15 v/v/v;  $n = 24$ ), of the  $A_{2a}$ -specific agonist CGS 21680 (0.5 mg kg<sup>-1</sup>;  $n = 15$ ), and of the nonspecific antagonist caffeine (25 mg kg<sup>-1</sup>;  $n = 16$ ). Similar effects were obtained for vertical activity. For drug treatments,  $P$  values indicate statistical significance when compared to solvent injection. For both drugs, a dose-response curve was determined on wild-type CD1 mice (not shown) and an active dose selected. **c**, Parameters reflecting increased anxiety in  $A_{2a}R^{-/-}$  mice. The relative frequency of entries into the open arms of the elevated plus-maze was recorded over a period of 5 min ( $n = 38$  and  $33$  for  $+/+$  and  $-/-$  mice, respectively). The relative time spent in the white compartment was measured over 5 min in the two-compartment test ( $n = 24$  and  $20$  for  $+/+$  and  $-/-$  mice, respectively). Error bars represent s.e.m.



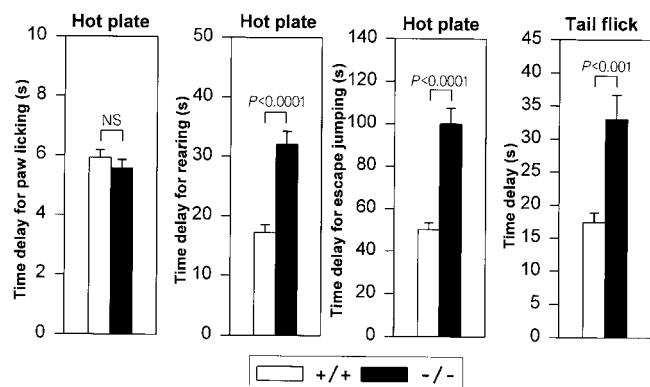
**Figure 4** Aggressive behaviour as measured by the resident-intruder test. Resident  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice ( $n = 20$  per group) were tested twice with a week interval (W1 and W2). Number of aggressed intruders over time, latency to first attack, and numbers of attacks and tail rattles over a 5-min period are represented. Animals that did not attack intruders were scored with 300 s. The Mann-Whitney  $U$ -test was used for attacks and tail rattles. Error bars represent s.e.m.

release of other pain mediators such as histamine and prostanoids by mast cells and basophils<sup>17</sup>. Reactivity to pain was investigated in knockout and control animals by using two tests involving thermal stimuli. In the hot-plate test, latencies for rearing and escape jumping were increased significantly, whereas licking latency was unaffected (Fig. 5). In the tail-flick test, mean latency was increased twofold for knockout mice. The higher nociceptive threshold suggests that the peripheral lack of  $A_{2a}$  receptor predominates over the spinal defect. This is in agreement with the observation that acute systemic administration of caffeine results in antinociception<sup>17</sup>.

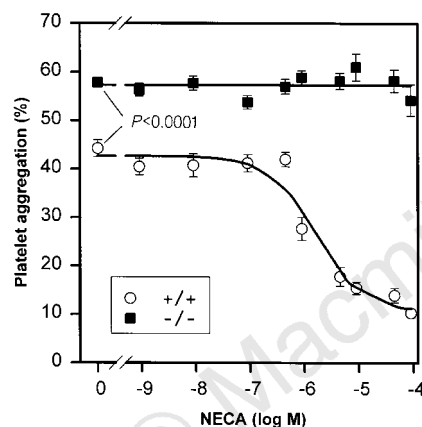
Adenosine and its analogues inhibit platelet aggregation by stimulating the formation of cyclic AMP, but the nature of the adenosine receptor involved in this action ( $A_{2a}$ , or a new subtype) has been called into question<sup>21</sup> on the basis of the low efficacy of  $A_{2a}$  agonists<sup>22</sup>. A count of blood cell populations, including platelets, showed no significant differences between  $A_{2a}R^{-/-}$  and  $A_{2a}R^{+/+}$  mice (data not shown). Platelet aggregation was assayed after addition of ADP, and changes in aggregation properties in response to the non-

selective agonist NECA were investigated. Aggregation was consistently more efficient in knockout animals (Fig. 6). NECA inhibited platelet aggregation fourfold in wild-type but not in knockout animals. The half-maximal effective concentration ( $EC_{50}$ ) of the anti-aggregation properties of NECA on mouse platelets was estimated to be 1.6  $\mu$ M, a value similar to that for human platelets (0.49  $\mu$ M)<sup>22</sup>. These data were confirmed with the  $A_{2a}$ -selective agonist CGS 21680 (not shown). The lack of effect of NECA and CGS 21680 on  $A_{2a}R^{-/-}$  platelet aggregation shows that the classical  $A_{2a}$  receptor is responsible for the anti-aggregation properties of adenosine and its analogues. The fact that N6-substituted adenosine derivatives, such as NECA and CGS 21680, bind to a low-affinity site known as the adenotin binding site, which is abundant on platelets, may explain why these agents are not very effective<sup>23</sup>. The increased aggregation of platelets from  $A_{2a}R^{-/-}$  mice indicates that the assay is affected by the adenosine released from control platelets.

As adenosine mediates cardiovascular effects, we determined blood pressure and heart rate in  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice, as well as their haemodynamic responses to CGS 21680. CGS 21680



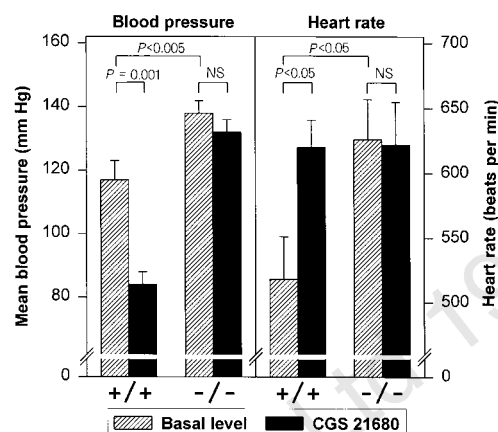
**Figure 5** Behavioural response to acute thermal stimuli. The latency for paw licking, rearing and escape jumping was measured in the hot-plate test for  $A_{2a}R^{+/+}$  ( $n = 46$ ) and  $A_{2a}R^{-/-}$  ( $n = 46$ ) mice. The reaction time was also measured in the tail-flick test ( $n = 15$  for both groups). Error bars represent s.e.m. NS, not significant.



**Figure 6** Platelet aggregation in  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice. Platelet-rich plasma was prepared and aggregation was measured in control conditions ( $n$ , 34 and 39 for  $A_{2a}R^{+/+}$  and  $A_{2a}R^{-/-}$  mice, respectively), and in the presence of the agonist NECA ( $n$ , was 3 to 8 for each concentration). Error bars represent s.e.m.

administration in wild-type mice resulted in immediate hypotension and reactive tachycardia (Fig. 7), in agreement with the  $A_{2a}R$ -mediated vasodilating effect of adenosine analogues on vascular beds<sup>1</sup>. The  $A_{2a}R$  gene deficiency caused a significant increase in blood pressure, suggesting that endogenous adenosine has a tonic vasodilating effect in wild-type mice. The heart rate was also increased (Fig. 7), which cannot be explained on the basis of the known physiological roles of adenosine on haemodynamics; this unexpected effect could result from complex compensatory mechanisms and may also contribute to the raised blood pressure. The hypotensive and tachycardic responses to CGS 21680 were abolished in  $A_{2a}R^{-/-}$  mice, confirming the participation of  $A_{2a}R$  in these actions (Fig. 7).

The  $A_{2a}$  receptor is a target for drug design: antagonists have been considered as treatments for Parkinson's disease, stroke, pain and inflammatory disorders<sup>24</sup>. But owing to a lack of selective drugs for different classes of adenosine receptors, several functions mediated by adenosine cannot be attributed to a specific receptor type. Our knockout model should enable the functions of the  $A_{2a}$  and other adenosine receptor types to be clarified and the potential therapeutic benefit of drugs acting on these receptors to be tested. Different features of the phenotype should help the identification of defects of the  $A_{2a}R$  gene in human disease. □



**Figure 7** Basal mean blood pressure and heart rate were measured on anaesthetized  $A_{2a}R^{+/+}$  ( $n = 9$ ) and  $A_{2a}R^{-/-}$  ( $n = 11$ ) mice by carotid artery catheterization. The measurements were repeated following bolus administration of the agonist CGS 21680 ( $1 \text{ mg kg}^{-1}$ ) through the catheter. Error bars represent s.e.m.

## Methods

**Drugs.** 2-[p-(2-carboxyethyl)phenethylamino]-5'-N-ethylcarboxamidoadenosine (CGS 21680), caffeine, and 5'-N-ethyl-carboxamidoadenosine (NECA) were all purchased from RBI.

**Targeting the  $A_{2a}R$  gene.** The  $A_{2a}R$  gene was cloned from a 129/Sv mouse genomic library by using a dog cDNA probe<sup>23</sup>. Seven clones spanning the locus were obtained, and the two exons were mapped and sequenced. A PGK-Neo cassette was inserted between the NcoI and EcoRI restriction sites within the first exon. The cassette replaced the first 102 codons of the gene, encoding transmembrane segments 1 to 3, and was flanked by 4.5 kb of 5' and 10.5 kb of 3' mouse gene fragments. Homologous recombination was carried out in the R1 ES cell line (129/Sv × 129/Sv-CP derived)<sup>25</sup>. ES cells were electroporated (BioRad Gene Pulser; 240 V and 500 μF) with 84 μg targeting construct DNA, and plated onto γ-irradiated STP OB500 feeder cells. G418(250 μg ml<sup>-1</sup>)-resistant colonies were collected after 11 days of selection, and recombinant clones were screened by Southern blotting after *Dra*I digestion and hybridization with a <sup>32</sup>P-labelled 2,800-bp *Eco*RI fragment (Fig. 1). After mild trypsinization, clumps of ES cells were allowed to aggregate overnight in M16 medium with single CD1 eight-cell stage embryos after removal of the zona pellucida by treatment with acidic Tyrode's buffer<sup>26</sup>. Embryos were transferred into the uterus of pseudopregnant recipients (2.5 d.p.c.). Fetuses were recovered by caesarean section on day 20. Chimaeras were mated with CD1 mice and heterozygous mutants were bred for four generations with wild-type CD1 mice in order to dilute out the 129/Sv background. Heterozygotes were then bred together to generate wild-type homozygotes ( $A_{2a}R^{+/+}$ ), heterozygotes ( $A_{2a}R^{+/-}$ ) and knockout ( $A_{2a}R^{-/-}$ ) animals. Heterozygotes have not so far been studied extensively.

**Morphology.** For routine light microscopy, sections of 10% formaldehyde-fixed, paraffin-embedded tissues were stained with haematoxylin and eosin. For *in situ* hybridization and binding assays, brains were frozen in 2-methylbutane and serially cut 15-μm-thick coronal sections were thaw-mounted onto poly-L-lysine- or gelatin-coated slides, respectively. Mouse oligonucleotide probes were as follows: 5'-GGCCAAGGTGTCTCCCTCATCTGCATCCTT TTTCATGAAG CCGCC-3' (proenkephalin A); 5'-AGGGCCACTT GTTTTTCAC TGAGGAATCA GCATCCCGCT TGCCC-3' (protachykinin); 5'-CTGTTGATCC ACACGGCCCA GCACACAAGC ACGTTA CCCAG GATG-3' ( $A_{2a}R$  5' probe); 5'-GGCCCTCGCC GCAGGTCTTT GTGGAGTCT CATCTGTCTG ACTGC-3' ( $A_{2a}R$  3' probe). Hybridization with <sup>35</sup>S-labelled oligonucleotides has been described<sup>9</sup>. Autoradiography was done with Hyperfilm β-max films (Amersham) for 3 d (proenkephalin) or 15 d (protachykinin,  $A_{2a}R$ ). The distributions of proenkephalin, protachykinin and  $A_{2a}R$  transcripts agreed with those reported for rat<sup>9</sup>. No signal was observed

after RNase pretreatment. For quantitative autoradiography, average optical densities (OD) over the dorsal striatum were measured on X-ray film autoradiographs using an IBAS image analyser (Zeiss Kontron). The measurement field was interactively defined and  $512 \times 512$  pixel, 256 grey-level images were generated using a MTI video camera with fixed black level and gain. The background level was subtracted for each section, and mean OD was determined for right and left striata. Five or more sections per mouse were analysed, and results were analysed by the Mann-Whitney *U*-test. Autoradiography binding assays have been described<sup>27</sup> and used [<sup>3</sup>H]CGS 21680 (NEN;  $48.6 \text{ Ci mmol}^{-1}$ ) as ligand and  $10 \mu\text{M}$  NECA as competitor.

**Blood pressure and heart rate.** Mice were anaesthetized using halothane inhalation (1–2% in oxygen). The left carotid artery was isolated and a PE-10 tubing catheter was inserted into the artery, secured with silk, and tunnelled subcutaneously to exit at the back of the neck. Mice were allowed to recover from surgery for at least 4 h. The arterial line was connected to a pressure transducer, and blood pressure and heart rate were measured using a computerized data acquisition system. For agonist study, a bolus of CGS 21680 ( $1 \text{ mg kg}^{-1}$ ) was administered through the arterial catheter.

**Platelet aggregation.** Blood was collected by cardiac puncture on sodium citrate (0.38%) and pooled from about 10 mice. Platelet-rich plasma was prepared by centrifugation and adjusted to  $1.75 \times 10^8 \text{ ml}^{-1}$ . Platelets were aggregated using a platelet-aggregation profiler (model PAP-4; Bio Data). Platelet-rich plasma ( $100 \mu\text{l}$ ) was incubated for 5 min in siliconized cuvettes at  $37^\circ\text{C}$ , and aggregation initiated by addition of  $5 \mu\text{M}$  ADP.

**Behavioural studies.** Male mice were kept on a 7 a.m.–7 p.m. light cycle. Exploratory activity (horizontal and vertical) was monitored in the  $20 \times 20 \text{ cm}$  arena of a Digiscan actimeter (Omnitech Electronics) during a 60-min period. An elevated plus-maze, with  $18 \times 6 \text{ cm}$  arms, was placed 50 cm above the floor. The mouse was placed at the centre of the maze, with head facing a closed arm. The number of entries, time spent and distance travelled in each arm and central area over a 5-min period was recorded by an automated image analysis system (Videotrack 512 system, Viewpoint). For the black and white compartments test, a 5-cm opening connected the compartments ( $32 \times 22 \text{ cm}$ , 18 cm height). Mice were placed in the black compartment (head facing a corner). The delay for first entering into the brightly lit white compartment, the time spent in the white compartment, and the number of entries were recorded for 5 min. The hot-plate test has been described<sup>28</sup>; the tail-flick test was done by using a tail-flick apparatus. Two measurements were made on each mouse at 3 and 5 cm from the tail tip. Aggressive behaviour was assayed by the resident-intruder test<sup>15</sup>, with slight modifications. Resident male  $A_{2A}R^{+/+}$  and  $A_{2A}R^{-/-}$  mice (14 weeks old) were isolated in cages ( $27 \times 13 \times 13 \text{ cm}$ ) without replacement of the litter for the four weeks preceding the test. 8-week-old outbred male CD1 mice, housed eight per cage, were used as intruders. A randomly chosen intruder, used only once in each session, was introduced in the resident cage, and aggressive behaviour recorded during a 5-min period.

**Statistical analysis.** Unless otherwise state, statistical analysis was done by the two-tailed unpaired Student's *t*-test (Instat program, GraphPad Software). The two-tailed non-parametric Mann-Whitney *U*-test was used when variances were not equal among series.

Received 21 April; accepted 5 June 1997.

- Olah, M. E. & Stiles, G. L. Adenosine receptor subtypes: characterization and therapeutic regulation. *Annu. Rev. Pharmacol. Toxicol.* **35**, 581–606 (1995).
- Libert, F. *et al.* Selective amplification and cloning of four new members of the G protein-coupled receptor family. *Science* **244**, 569–572 (1989).
- Maenhaut, C. *et al.* RDC8 codes for an adenosine  $A_2$  receptor with physiological constitutive activity. *Biochem. Biophys. Res. Commun.* **173**, 1169–1178 (1990).
- Fredholm, B. B. Adenosine, adenosine receptors and the actions of caffeine. *Pharmacol. Toxicol.* **76**, 93–101 (1995).
- Fink, J. S. *et al.* Molecular cloning of the rat  $A_2$  adenosine receptor: selective co-expression with  $D_2$  dopamine receptors in rat striatum. *Mol. Brain Res.* **14**, 186–195 (1992).
- Furlong, T. J., Pierce, K. D., Selbie, L. A. & Shine, J. Molecular characterization of a human brain adenosine  $A_2$  receptor. *Mol. Brain Res.* **15**, 62–66 (1992).
- Weaver, D. R.  $A_{2A}$  adenosine receptor gene expression in developing rat brain. *Mol. Brain Res.* **20**, 313–327 (1993).
- Höpker, V. H., Saffrey, M. J. & Burnstock, G. The neurotogenic effect of myenteric plexus on striatal neurones in co-culture involves nitric oxide. *Neuroreport* **6**, 1153–1156 (1995).
- Schiffmann, S. N. & Vanderhaeghen, J. J. Adenosine  $A_2$  receptors regulate the gene expression of striatopallidal and striatonigral neurons. *J. Neurosci.* **13**, 1080–1087 (1993).
- Balk, J. H. *et al.* Parkinsonian-like locomotor impairment in mice lacking dopamine  $D_2$  receptors. *Nature* **377**, 424–428 (1995).
- Grace, A. A. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience* **41**, 1–24 (1991).
- Nikodijević, O., Jacobson, K. A. & Daly, J. W. Locomotor activity in mice during chronic treatment

- with caffeine and withdrawal. *Pharmacol. Biochem. Behav.* **44**, 199–216 (1993).
- File, S. E., Baldwin, H. A., Johnston, A. L. & Wilks, L. J. Behavioral effects of acute and chronic administration of caffeine in the rat. *Pharmacol. Biochem. Behav.* **30**, 809–815 (1988).
- Jain, N., Kemp, N., Adeyemo, O., Buchanan, P. & Stone, T. W. Anxiolytic activity of adenosine receptor activation in mice. *Br. J. Pharmacol.* **116**, 2127–2133 (1995).
- Saudou, F. *et al.* Enhanced aggressive behavior in mice lacking 5-HT<sub>1B</sub> receptor. *Science* **265**, 1875–1878 (1994).
- Palmour, R. M., Lipowski, C. J., Simon, C. K. & Ervin, F. R. Adenosine analogs inhibit fighting in isolated male mice. *Life Sci.* **44**, 1293–1301 (1989).
- Sawynok, J. & Yaksh, T. L. Caffeine as an analgesic adjuvant: a review of pharmacology and mechanisms of action. *Pharmacol. Rev.* **45**, 43–85 (1993).
- Sweeney, M. L., White, T. D., Jhamandas, K. H. & Sawynok, J. Morphine releases endogenous adenosine from the spinal cord in vivo. *Eur. J. Pharmacol.* **141**, 169–170 (1987).
- De Lander, G. E. & Keil, G. J. Antinociception induced by intrathecal coadministration of selective adenosine receptor and selective opioid receptor agonists in mice. *J. Pharmacol. Exp. Ther.* **268**, 943–951 (1994).
- McQueen, D. S. & Ribeiro, J. A. Pharmacological characterization of the receptor involved in chemocentric induction by adenosine. *Br. J. Pharmacol.* **88**, 615–620 (1986).
- Cristalli, G. *et al.* Inhibition of platelet aggregation by adenosine receptor agonists. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **349**, 644–650 (1994).
- Dionisotti, S., Zocchi, C., Varani, K., Borea, P. A. & Ongini, E. Effects of adenosine derivatives on human and rabbit platelet aggregation. Correlation of adenosine receptor affinities and antiaggregatory activity. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **346**, 673–676 (1994).
- Varani, K., Gessi, S., Dalpiaz, A. & Borea, P. A. Pharmacological and biochemical characterization of purified  $A_{2A}$  adenosine receptors in human platelet membranes by [<sup>3</sup>H]-CGS 21680 binding. *Br. J. Pharmacol.* **117**, 1693–1701 (1996).
- Ongini, E. & Fredholm, B. B. Pharmacology of adenosine  $A_{2A}$  receptors. *Trends Pharmacol. Sci.* **17**, 364–372 (1996).
- Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
- Worst, W. & Joyner, A. L. in *Gene Targeting, a Practical Approach* (ed. Joyner, A. L.) 33–61 (Oxford University Press, Oxford, 1993).
- Schiffmann, S. N. & Vanderhaeghen, J. J. Ontogeny of gene expression of adenosine  $A_2$  receptor in the striatum: early localization in the patch compartment. *J. Comp. Neurol.* **317**, 117–128 (1992).
- Costentin, J. *et al.* Dissociated effects of inhibitors of enkephalin-metabolising peptidases or naloxone on various nociceptive responses. *Eur. J. Pharmacol.* **123**, 37–44 (1986).

**Acknowledgements.** The R1 ES and STP cell lines were provided by A. Nagy and B. Colledge, respectively. We thank J.-F. Aubert for his help with the determination of the haemodynamic parameters; O. Le Moine, H. Louis, D. Penninck and C. Kucharszewski for discussion; and E. Bressy, M. J. Simons, E. Quertainmont and M. Verslype for technical assistance. This work was supported by the Belgian programme on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Federal Service for Science, Technology and Culture. It was also supported by the Fonds de la Recherche Scientifique Médicale of Belgium, the BIOMED2 programme, and the Fondation Médicale Reine Elisabeth. Scientific responsibility belongs to the authors. C.L. and S.N.S. are Chercheurs Qualifiés of the Fonds National de la Recherche Scientifique.

Correspondence and request for materials should be addressed to M.P. (e-mail: mparment@ulb.ac.be). The EMBL/Genbank accession numbers for the mouse  $A_{2A}$ R gene sequences are Y13344, Y13345 and Y13346.

## Altered pain perception and inflammatory response in mice lacking prostacyclin receptor

Takahiko Murata\*, Fumitaka Ushikubi\*, Toshiyuki Matsuoka\*, Masakazu Hirata\*, Atsushi Yamasaki†, Yukihiko Sugimoto†, Atsushi Ichikawa†, Yoshiya Aze‡, Takashi Tanaka§, Nobuaki Yoshida§, Akinori Ueno||, Sachiko Oh-ishi|| & Shuh Narumiya\*

\* Department of Pharmacology, Faculty of Medicine, and † Department of Physiological Chemistry, Faculty of Pharmaceutical Sciences, Kyoto University, Sakyo, Kyoto 606-01, Japan

‡ Fukui Institute for Safety Research, Ono Pharmaceutical Company, Mikuni, Fukui 913, Japan

§ Division of Molecular and Cellular Immunology, Research Institute Osaka Medical Center for Maternal and Child Health, Izumi, Osaka 590-02, Japan

|| Department of Pharmacology, School of Pharmaceutical Sciences, Kitasato University, Shirokane, Tokyo 108, Japan

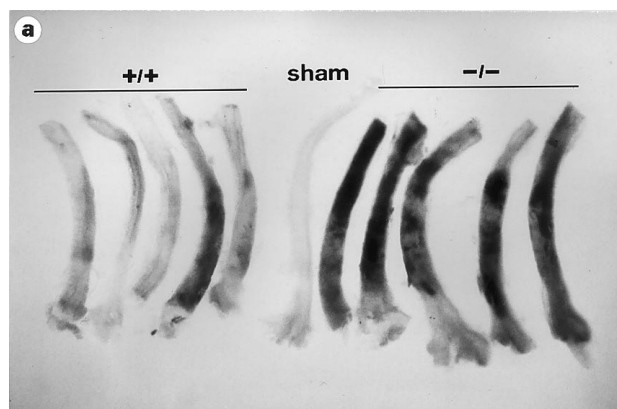
Prostanoids are a group of bioactive lipids working as local mediators<sup>1</sup> and include D, E, F and I types of prostaglandins (PGs) and thromboxanes. Prostacyclin (PGI<sub>2</sub>) acts on platelets and blood vessels to inhibit platelet aggregation and to cause vasodilatation, and is thought to be important for vascular homeostasis<sup>2</sup>. Aspirin-like drugs, including indomethacin,



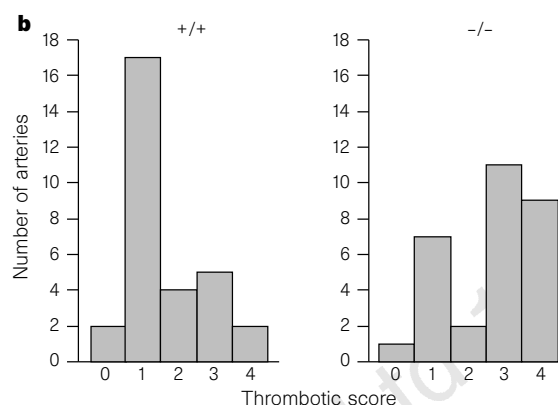
We disrupted the mouse gene encoding the prostacyclin receptor (*IP*) (Fig. 1a) and established two lines of embryonic stem cells.

Chimaeric mice were generated from each line. The mice were backcrossed with C57BL/6 mice and the resulting heterozygous littermates ( $IP^{+/-}$ ) were bred to produce homozygous mice (Fig. 1b). Breeding of heterozygotes produced homozygous  $IP^{-/-}$  mice at a rate lower than expected from the mendelian frequency; 30 (+/+), 72 (+/-) and 19 (-/-) males and 48 (+/+), 80 (+/-) and 32 (-/-) females. This suggests that there may be prenatal mortality in  $IP^{-/-}$  deficient mice, especially in males. A similar abnormal mendelian ratio of male  $F_2$  progeny was reported for cyclooxygenase-2 deficient mice<sup>6</sup>. Homozygous mice that survived to birth grew normally, lived longer than one year, and were fertile. No morphological or histological abnormalities were detected. We analysed one line of  $IP$  mutant mice in detail.

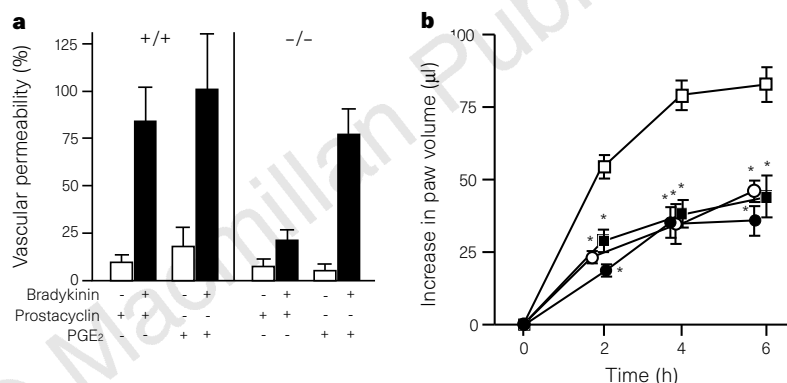
679



**Figure 2** Arterial thrombosis in wild-type and IP-deficient homozygous mice. **a**, Thrombosis induced by ferric chloride. Five arterial preparations each from wild-type (left) and  $IP^{-/-}$  mice (right) are shown. Middle, preparation from a sham-operated homozygous animal. **b**, Thrombotic scores in wild-type (+/+) (left) and  $IP^{-/-}$  (right) mice. Thrombosis was scored as follows: 0, without apparent



thrombus; 1, small and isolated thrombi; 2, mural thrombi; 3, partially occlusive thrombi; 4, occlusive thrombi. Results from fifteen mice (a total of 30 arteries) are shown for each. There is a significant difference at  $P < 0.01$  between wild-type and  $IP^{-/-}$  mice by the Mann-Whitney  $U$ -test.



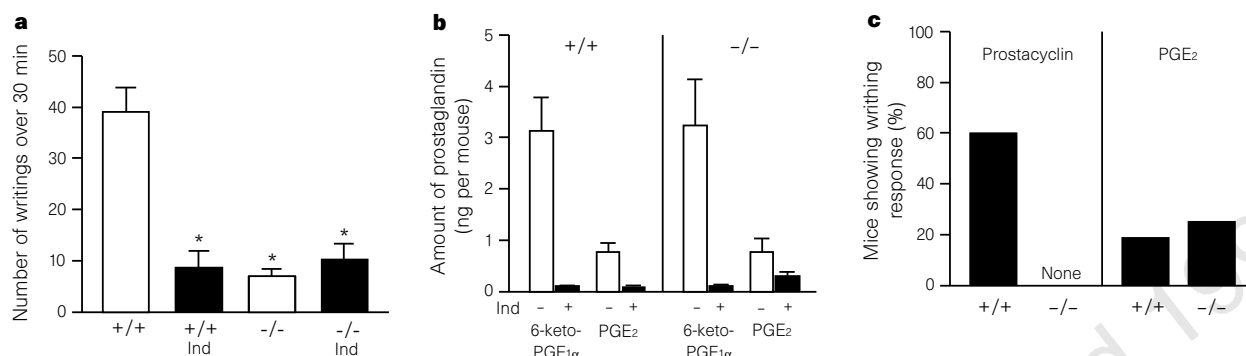
**Figure 3** Inflammatory responses in IP-deficient mice. **a**, Vascular permeability to exogenous prostacyclin and  $PGE_2$  in wild-type and  $IP^{-/-}$  mice. Either prostacyclin ( $n = 6$ ) or  $PGE_2$  ( $n = 5$ ) (1 nmol of each) was injected intradermally alone or with 0.1 nmol bradykinin. Vascular permeability was examined as described in Methods. **b**, Carrageenin-induced paw oedema. 30  $\mu$ l of 2% carrageenin was

injected into the right hind paw of mice and the increase in paw volume was monitored. Indomethacin (Ind) (10 mg kg<sup>-1</sup>) was administered intraperitoneally 30 min before carrageenin treatment. White squares, wild-type ( $n = 9$ ); white circles,  $IP^{-/-}$  mice ( $n = 10$ ); black squares, wild-type mice treated with indomethacin ( $n = 8$ ); black circles,  $IP^{-/-}$  mice treated with indomethacin ( $n = 7$ ). \*,  $P < 0.01$ .

rings isolated from IP-deficient homozygous mice did not relax in response to cicaprost, although the aorta from wild-type mice did (Fig. 1d). Furthermore, intravenous injection of cicaprost (1  $\mu$ g kg<sup>-1</sup>) caused hypotension of ~30 mm Hg in anaesthetized wild-type mice, whereas there was no change in blood pressure in IP-deficient mice even at a cicaprost dose of 10  $\mu$ g kg<sup>-1</sup> (Fig. 1e). On the other hand, intravenous administration of the prostaglandin  $PGE_2$  caused transient hypotension in both types of animals; a drop of  $24.0 \pm 2.8$  ( $n = 5$ ) and  $23.2 \pm 1.7$  mm Hg ( $n = 5$ ) at 10  $\mu$ g kg<sup>-1</sup> of  $PGE_2$  for wild-type and IP-deficient mice, respectively. These results verified that specific disruption of the IP receptor had occurred, and further indicate that a single type of IP receptor mediates the actions of prostacyclin in both platelets and vascular smooth muscles. The basal blood pressure and heart rate of IP-deficient mice were identical with those of wild-type mice in the conscious state: the mean blood pressure and heart rate were  $88 \pm 3$  mm Hg and  $646 \pm 23$  beats per min ( $n = 10$ ), and  $85 \pm 3$  mm Hg and  $645 \pm 21$  beats per min ( $n = 10$ ) for wild-type and IP-deficient mice, respectively. The bleeding time in IP-deficient mice was normal, being  $87 \pm 10$  and  $90 \pm 11$  s for wild-type and  $IP^{-/-}$  mice, respectively ( $n = 12$  each). However, when a prostacyclin mimetic, beraprost<sup>8</sup>, was administered *per os* (3 mg kg<sup>-1</sup>),

it prolonged the bleeding time of wild-type mice (>10 min;  $n = 6$ ) but not that of IP-deficient mice ( $117 \pm 14$  s;  $n = 6$ ). These results suggest that prostacyclin is not involved in the regulation of heart rate, blood pressure and bleeding under basal conditions. This is in contrast with the findings on another endothelium-derived vasorelaxant, nitric oxide; mice deficient in endothelial nitric oxide synthase are hypertensive under basal conditions<sup>9</sup>.

Endogenous prostacyclin has long been thought to be involved in the prevention of thrombosis and vasospasm<sup>2</sup>. We therefore evaluated the thrombotic tendency of IP-deficient mice. We first confirmed that  $IP^{-/-}$  mice could synthesize prostacyclin. Isolated aortas from  $IP^{-/-}$  mice produced  $126 \pm 22$  pmol ( $n = 5$ ) of prostacyclin, determined as the stable metabolite 6-keto-PGF<sub>1 $\alpha$</sub>  that was produced per mg wet weight in 20 min. These levels were similar to those of wild-type mice ( $177 \pm 40$  pmol;  $n = 5$ ). The bilateral carotid arteries of wild-type and  $IP^{-/-}$  mice were then exposed and topically painted with 7.5% FeCl<sub>3</sub> solution, a procedure known to injure the endothelium<sup>10</sup>. Occlusion of the arterial flow and thrombus formation were examined after 4 h. As shown in Fig. 2a, most of the arteries of wild-type mice showed only mural thrombi of yellowish colour. In contrast, about two-thirds showed obstructive



**Figure 4** Nociceptive responses in IP-deficient mice. **a**, Acetic acid writhing. Acetic acid was injected intraperitoneally into wild-type mice ( $n = 8$ ), wild-type mice treated with indomethacin (Ind) ( $n = 8$ ),  $IP^{-/-}$  mice ( $n = 9$ ) and  $IP^{-/-}$  mice treated with indomethacin ( $n = 6$ ), and the writhing responses were counted for 30 min. Indomethacin was administered as for Fig. 3. \*,  $P < 0.01$ . **b**, Production of prostacyclin and PGE<sub>2</sub> in the peritoneal cavity of mice treated with acetic acid. Diluted acetic acid was injected i.p. and the peritoneal fluid collected after 15 min

for measuring PGs. Results are means  $\pm$  s.e.m. ( $n$  was 4–5). Ind, indomethacin-pretreatment. **c**, Writhing responses to exogenous prostacyclin and PGE<sub>2</sub>. 2  $\mu$ g of each was administered intraperitoneally to wild-type (+/+) and IP-deficient (-/-) mice, and the writhing response was counted in affected animals for 15 min. The numbers of animals used were 20 for prostacyclin injection and 16 for PGE<sub>2</sub> injection.

thrombi with reddish tails (in  $IP^{-/-}$  mice) (Fig. 2b). In another experiment monitoring the effects of the FeCl<sub>3</sub> treatment for a longer period, 30% of IP-deficient mice (4 out of 12) died within 1 day owing to bilateral occlusions of the carotid arteries and/or embolic stroke, whereas the wild-type mice ( $n = 13$ ) all survived this period.

Acute inflammation is characterized by hyperaemia, oedema, leukocyte migration and pain. Prostanoids are thought to contribute to these processes by causing peripheral vasodilation and by sensitizing primary afferents to noxious stimuli<sup>3</sup>. PGE<sub>2</sub> has long been regarded as a principal mediator of these processes, although several prostanoids show similar effects<sup>11,12</sup>. We therefore first compared the effects of exogenously added PGE<sub>2</sub> and prostacyclin on vascular permeability. As shown in Fig. 3a, both PGE<sub>2</sub> and prostacyclin potentiated the bradykinin-induced permeability to a similar extent in wild-type mice;  $IP^{-/-}$  mice showed a response to PGE<sub>2</sub> only. We then examined the role of endogenous prostacyclin in inflammation by applying the carrageenin-induced paw-oedema model<sup>13</sup> to  $IP^{-/-}$  mice. The carrageenin injection caused a time-dependent increase in the volume of the injected paw in wild-type animals as a result of oedema (Fig. 3b). This increase is suppressed by 50% by pretreatment with indomethacin. Unexpectedly, the progress of oedema was reduced in IP-deficient mice to the same extent as in indomethacin-treated animals, indicating that prostacyclin is a major prostanoid eliciting the inflammatory response in this model. The proinflammatory role of prostacyclin was also tested in another model, carrageenin-induced pleurisy<sup>14</sup>, in which carrageenin was injected into the pleural cavity of mice and exudate was collected after 3 h. A significant reduction of the exudate volume was observed in IP-deficient mice ( $P < 0.05$ ) (data not shown).

The role of prostacyclin in inflammatory pain was next examined in IP-deficient mice. Our interest in the involvement of IP in pain originates from our findings that IP receptor messenger RNA is abundantly expressed and coexists with substance-P mRNA in neurons of the dorsal root ganglion<sup>15</sup>. Prostanoids may produce hyperalgesia not only by peripheral sensitization<sup>3</sup> but also by facilitating pain transmission at the spinal<sup>16</sup> and supraspinal<sup>17</sup> levels. IP involvement in the latter mechanisms was tested by the tail-flick method<sup>18,19</sup> and by the hot-plate test<sup>18</sup>. The latent period before response in both tests was not significantly different in wild-type and IP-deficient mice (data not shown), suggesting that the receptor IP is not involved in nociceptive transmission at the spinal and supraspinal levels. We next used the acetic acid-induced wri-

thing test<sup>20</sup> to examine the involvement of IP in inflammatory pain. Injection of a dilute acetic acid solution intraperitoneally caused about forty writhing responses in wild-type animals; this reaction was significantly reduced by pretreatment with indomethacin (Fig. 4a). Surprisingly, the response of the IP-deficient mice was comparable to that of the indomethacin-pretreated animals and markedly reduced compared to wild-type mice, indicating that IP is involved in this nociceptive response. Under these conditions, acetic acid injection produced similar amounts of prostacyclin in wild-type and  $IP^{-/-}$  mice (Fig. 4b). PGE<sub>2</sub> was also produced, but in about one quarter of the amount of prostacyclin. The same profile for PG production has been reported in the peritoneal cavity of zymosan-injected mice<sup>21</sup>. Because PGE<sub>2</sub> may be involved in pain response, we compared the writhing responses to prostacyclin and PGE<sub>2</sub> by injecting the same amount of each PG intraperitoneally into these animals (Fig. 4c). Prostacyclin caused the writhing response in 60% of the wild-type mice and in none of the IP-deficient animals, whereas PGE<sub>2</sub> induced the response in less than 25% animals in both types.

In conclusion, we have established the antithrombotic role of endogenous prostacyclin. As prostacyclin has also been implicated in atherosclerosis and diabetic angiopathy<sup>2,22</sup>, IP-deficient mice may be useful for examining its role in these diseases. Our results indicate that prostacyclin is a mediator of inflammatory swelling and pain. Whether prostacyclin and PGE<sub>2</sub> have different, context-dependent roles in these processes is an important issue that could be tested by comparing the response to various stimuli at different sites in IP-deficient mice and in mice deficient in PGE receptors. □

## Methods

**Gene targeting.** Murine IP genomic clones were isolated from a 129/Sv-strain genomic DNA library (Stratagene) by using its cDNA<sup>3</sup> as a probe. The targeting vector was constructed by replacing a 2.4-kb fragment containing parts of the putative exons II and III, which encode the sixth transmembrane domain towards the C terminus of this seven-transmembrane-domain receptor<sup>5</sup>, with the neomycin-resistance gene (pMC1-neo; Stratagene). The herpes simplex virus thymidine kinase gene (TK) was inserted downstream (Fig. 1a). The targeting vector was linearized with Asp718 and introduced into E14-1 ES cells by electroporation. G418- and gancyclovir-resistant clones were isolated and screened by PCR for homologous recombination, which was then confirmed by Southern blot hybridization. The generation of chimaeras and mutant mice has been described<sup>23</sup>.

**Bioassays.** Blood was drawn from the heart and mixed with an equal volume of 20 mM HEPES, pH 7.4. Sodium citrate (0.38%) was used as anticoagulant.

Platelet-rich plasma was obtained by centrifugation, and diluted to  $5 \times 10^5$  platelets per  $\mu\text{l}$  with plasma. Platelet aggregation was assayed as described<sup>7</sup>. An aortic ring 4 mm long was prepared from thoracic aorta and mounted on a pair of wires in an organ bath filled with Krebs–Henseleit buffer gassed with 95%  $\text{O}_2$ –5%  $\text{CO}_2$  at 37 °C. Preload was 0.5 g. The ring was contracted with 0.1  $\mu\text{M}$  noradrenaline and its relaxation response to various concentrations of cicaprost tested.

**Blood pressure, heart rate and bleeding time.** Basal blood pressure and heart rate were measured in conscious animals by using the tail-cuff method. In the analysis of cicaprost- or  $\text{PGE}_2$ -induced hypotension, animals were anaesthetized with pentobarbital sodium (60 mg  $\text{kg}^{-1}$ ; i.p.), and arterial blood pressure was monitored through a cannula into the carotid artery. Bleeding time was measured as described<sup>24</sup>.

**Prostanoid production.** Isolated aorta was preincubated in oxygenated Krebs–Henseleit buffer at 37 °C for 1 h and then incubated in fresh buffer for 20 min; the supernatant was assayed. Prostanoids in peritoneal lavage were extracted with Sep-Pak C18 columns and then assayed for 6-keto-PGF $_{1\alpha}$  and  $\text{PGE}_2$  (ref. 25).

**Analysis of thrombus formation.** Mice (24–30 g) were anaesthetized with halothane and the carotid arteries exposed by a cervical incision. A piece of Parafilm 1 mm wide was inserted under the artery and 5  $\mu\text{l}$  of a 7.5%  $\text{FeCl}_3$  solution was dropped on the artery<sup>10</sup>. After 1 min the solution was wiped off, and the mice were allowed to recover from the anaesthetic. After 4 h, the arteries and blood flow were examined microscopically by cutting the distal end of the artery. A 5-mm long segment containing the affected region was excised and fixed with 10% formaldehyde. Thrombosis-induced death of animals treated with 5%  $\text{FeCl}_3$  solution was recorded after 24 h.

**Inflammation.** Vascular permeability was tested using Pontamine skyblue as described<sup>26</sup>. PGs alone, bradykinin alone, or both were intradermally injected. After 40 min, mice were killed. The exuded dye in the skin was extracted and the amounts determined. Permeability was expressed as a percentage of that induced by 10 nmol bradykinin alone in the same animals. Carrageenin-induced pleurisy was evoked by injecting 40  $\mu\text{l}$  of 2% carrageenin into the right pleural cavity of mice<sup>14</sup>. After 3 h, they were killed and the exudate volume measured.

**Nociceptive tests.** The tail-flick response was evoked either by heat<sup>18</sup> or by applying pressure<sup>19</sup>. Latencies were determined by using a tail-flick apparatus (Ugo Basile) and with an analgesy meter (Ugo Basile), respectively. In the hot-plate test, mice were placed on a plate ( $55 \pm 0.1$  °C) and the latency for jumping determined as described<sup>18</sup>. In the writhing test<sup>20</sup>, mice were intraperitoneally injected with either 0.9% acetic acid (5 ml  $\text{kg}^{-1}$ ) or PGs (2  $\mu\text{g}$ ). The frequency of stretch responses was counted for 30 min for acetic acid or for 15 min for PGs. Indomethacin (10 mg  $\text{kg}^{-1}$ ) was administered intraperitoneally 30 min before acetic acid injection.

**Data analysis.** Data are expressed as means  $\pm$  s.e.m.. The significance of difference between groups was evaluated using ANOVA with a subsequent Dunnett's test, except for thrombus formation (Fig. 2 legend).

Received 28 February; accepted 4 June 1997.

- Campbell, W. B. & Halushka, P. V. In *The Pharmacological Basis of Therapeutics* (9th edn) (eds Hardman, J. G. et al.) 601–616 (McGraw-Hill, New York, 1996).
- Bunting, S., Moncada, S. & Vane, J. R. The prostacyclin-thromboxane  $\text{A}_2$  balance: pathophysiological and therapeutic implications. *Br. Med. Bull.* **39**, 271–276 (1983).
- Davies, P., Bailey, P. J., Goldenberg, M. M. & Ford-Hutchinson, A. W. The role of arachidonic acid oxygenation products in pain and inflammation. *Annu. Rev. Immunol.* **2**, 335–357 (1984).
- Ushikubi, F., Hirata, M. & Narumiya, S. Molecular biology of prostanoid receptors: an overview. *J. Lip. Mediat.* **12**, 343–359 (1995).
- Namba, T. et al. cDNA cloning of a mouse prostacyclin receptor. *J. Biol. Chem.* **269**, 9986–9992 (1994).
- Morham, S. G. et al. Prostaglandin synthase 2 gene disruption causes severe renal pathology in the mouse. *Cell* **83**, 473–482 (1995).
- Armstrong, R. A. et al. Functional and ligand binding studies suggest heterogeneity of platelet prostacyclin receptor. *Br. J. Pharmacol.* **97**, 657–668 (1989).
- Murata, T. et al. General pharmacology of beraprost sodium: Effect on the autonomic, cardiovascular and gastrointestinal systems, and other effects. *Arzneim. Forsch.* **39**, 867–876 (1989).
- Huang, P. L. et al. Hypertension in mice lacking the gene for endothelial nitric oxide synthase. *Nature* **377**, 239–242 (1995).
- Kurz, K. D., Main, B. W. & Sandusky, G. E. Rat model of arterial thrombosis induced by ferric chloride. *Thromb. Res.* **60**, 269–280 (1990).
- Vadas, P., Wasi, S., Movat, H. Z. & Hay, J. B. Extracellular phospholipase  $\text{A}_2$  mediates inflammatory hyperaemia. *Nature* **293**, 583–585 (1981).
- Ferreira, S. H. Inflammatory pain, prostaglandin hyperalgesia and the development of peripheral analgesics. *Trends Pharmacol. Sci.* **2**, 183–186 (1981).
- Winter, C. A. & Flataker, L. Reaction thresholds to pressure in edematous hindpaws of rats and responses to analgesic drugs. *J. Pharmacol. Exp. Ther.* **150**, 165–171 (1965).

- Utsunomiya, I., Nagai, S. & Oh-ishi, S. Differential effects of indomethacin and dexamethasone on cytokine production in carrageenin-induced rat pleurisy. *Eur. J. Pharmacol.* **252**, 213–218 (1994).
- Oida, H. et al. *In situ* hybridization studies of prostacyclin receptor mRNA expression in various mouse organs. *Br. J. Pharmacol.* **116**, 2828–2837 (1995).
- Malmberg, A. B. & Yaksh, T. L. Hyperalgesia mediated by spinal glutamate or substance P receptor blocked by spinal cyclooxygenase inhibition. *Science* **257**, 1276–1279 (1992).
- Oka, T., Aou, S. & Hori, T. Intracerebroventricular injection of interleukin-1 $\beta$  induces hyperalgesia in rats. *Brain Res.* **624**, 61–68 (1993).
- Pierce, M. F. & Schroeder, L. A. Spinal and supraspinal sites for morphine and nefopam analgesia in the mouse. *Eur. J. Pharmacol.* **74**, 135–140 (1981).
- Millan, M. J. & Colpaert, F. C. 5-Hydroxytryptamine (5-HT) $_{1A}$  receptors and the tail-flick response. II. High efficacy 5-HT $_{1A}$  agonists attenuate morphine-induced antinociception in mice in a competitive-like manner. *J. Pharmacol. Exp. Ther.* **256**, 983–992 (1991).
- Koster, R., Anderson, M. & deBeer, E. J. Acetic acid for analgesic screening. *Fed. Proc.* **18**, 412 (1959). **FINAL PAGE?**
- Doherty, N. S. et al. The role of prostaglandins in the nociceptive response induced by intraperitoneal injection of zymosan in mice. *Br. J. Pharmacol.* **91**, 39–47 (1987).
- Vane, J. R., Anggard, E. E. & Botting, R. M. Regulatory functions of the vascular endothelium. *New Eng. J. Med.* **323**, 27–36 (1990).
- Kawabe, T. et al. The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. *Immunity* **1**, 167–178 (1994).
- Dejana, E., Callioni, A., Quintana, A. & Gaetano, G. Bleeding time in laboratory animals. II—A comparison of different assay conditions in rats. *Thromb. Res.* **15**, 191–197 (1979).
- Ritter, J. M., Aksoy, A., Cragoe, E. J. & Taylor, G. W. Actions of amiloride analogues on prostacyclin synthesis by rat aortic rings. *Br. J. Pharmacol.* **92**, 857–862 (1987).
- Ueno, A., Tokumasa, T., Naraba, H. & Oh-ishi, S. Involvement of bradykinin in endotoxin-induced vascular permeability increase in the skin of rats. *Eur. J. Pharmacol.* **284**, 211–214 (1995).

**Acknowledgements.** We thank K. Ishikawa and Y. Kataoka for ES cell injection and K. Okuyama for secretarial assistance. This work was supported by grants from the Ministry of Education, Science and Culture of Japan, the Japanese Society for Promotion of Sciences, the Uehara Memorial Foundation, the Smoking Research Foundation and the Japanese Foundation on Metabolism and Diseases.

Correspondence and requests for materials should be addressed to S.N. (e-mail: snaru@mfour.med.kyoto-u.ac.jp).

## Crustacean appendage evolution associated with changes in Hox gene expression

Michalis Averof\* & Nipam H. Patel†

Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK

† HHMI, University of Chicago, MC1028, N-101, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA

\* Present address: EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Homeotic (Hox) genes specify the differential identity of segments along the body axis of insects. Changes in the segmental organization of arthropod bodies may therefore be driven by changes in the function of Hox genes<sup>1–3</sup>, but so far this has been difficult to demonstrate. We show here that changes in the expression pattern of the Hox genes *Ubx* and *AbdA* in different crustaceans correlate well with the modification of their anterior thoracic limbs into feeding appendages (maxillipeds). Our observations provide direct evidence that major morphological changes in arthropod body plans are associated with changes in Hox gene regulation. They suggest that homeotic changes<sup>1,4</sup> may play a role in the normal process of adaptive evolutionary change.

Genetic manipulation of model organisms like *Drosophila* can reveal the potential of developmental systems to undergo particular types of morphological change: for example, patterns of segmental specialization may be altered by changing the function of Hox genes. This approach alone, however, cannot identify the actual genetic changes that take place over macro-evolutionary timescales. Comparative studies are required but these so far indicate that crustaceans and insects share almost identical complements of Hox genes<sup>5</sup> and that the domains of Hox gene expression have been broadly conserved during insect evolution (reviewed in ref. 6). Initial comparisons of Hox gene expression between crustaceans and insects documented changes associated with the independent specialization of trunk regions (thorax and abdomen) in these



animals<sup>7,8</sup>. Studies on such divergent animals are constrained by the lack of a common framework for morphological comparison. We focus here on comparisons between different subgroups of crustaceans, where we can identify changes in Hox gene expression that relate to well characterized variations in patterns of segmental specialization.

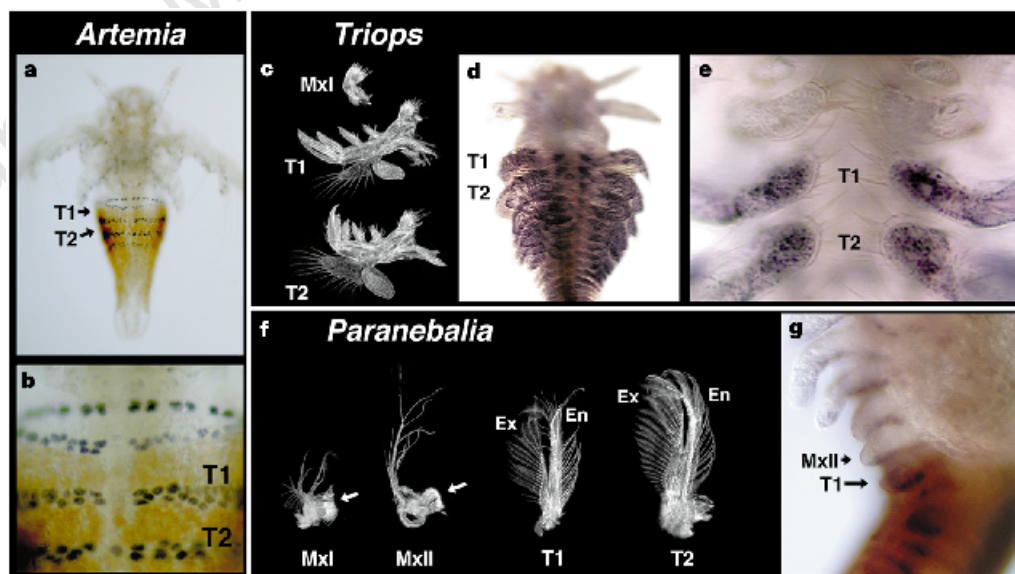
In most arthropods, thoracic appendages are specialized primarily for locomotor functions. But in some crustaceans, limbs from the anterior thorax have been recruited and specialized for the manipulation of food. These modified thoracic appendages, termed maxillipeds, are morphologically and functionally more similar to the feeding appendages (maxillae) of the more anterior gnathal segments than to the remaining thoracic limbs. Maxillipeds are generally reduced in size, show modification, fusion, or loss of particular limb elements, and are primarily associated with feeding, not locomotion<sup>9–12</sup>. These morphological specializations are reminiscent of partial homeotic transformations<sup>4</sup> (thoracic to gnathal) and represent the evolution of new and distinct segmental identities in the anterior region of the trunk. We investigated whether any of these changes could be associated with differences in Hox gene regulation by studying the combined expression of the Hox genes *Ubx* and *abdA* in diverse crustacean species. In crustaceans, these genes are thought to be involved in specifying the post-gnathal region of the trunk<sup>7</sup>. Using the monoclonal antibody FP6.87 (ref. 13), which recognizes a conserved epitope specific to the *Ubx* and *AbdA* proteins<sup>7,13</sup>, we examined *Ubx*–*AbdA* expression in thirteen crustacean species from nine different orders. In all cases, we observed expression in the limb-bearing region of the trunk posterior to gnathal segments, or posterior to thoracic segments bearing maxillipeds when these were present.

In branchiopod crustaceans, which do not have maxillipeds, *Ubx* and *abdA* are expressed throughout the thoracic region. In *Artemia franciscana* (Anostraca), there is FP6.87 staining in a domain that starts from the first thoracic segment (T1) and extends posteriorly

throughout the thorax<sup>7</sup> (Fig. 1a, b). This corresponds to a region of the body where all segments and the associated limbs develop similar morphological characteristics<sup>7</sup>. Expression patterns are similar in a related anostracan species (fairy shrimp; data not shown). In *Triops longicaudatus* (Notostraca), the same anterior boundary of *Ubx*–*AbdA* distribution is seen in the early larval stages, at a time when all thoracic limbs are morphologically identical and clearly distinct from the gnathal appendages (Fig. 1c–e).

A similar pattern of expression is also observed in leptostracans, a group of malacostracan crustaceans that bear no maxillipeds in anterior trunk segments. The thorax of all malacostracans consists characteristically of eight segments; leptostracans are thought to have retained a number of primitive malacostracan features, including uniform morphology of all eight thoracic segments and limbs<sup>9</sup> (Fig. 1f). In early embryos of *Paranebalia belizensis* (Leptostraca), we observed FP6.87 staining in the entire post-gnathal region of the trunk, starting from T1 (Fig. 1g). We consider this anterior boundary of *Ubx*–*abdA* expression to be primitive within malacostracans and shared by all crustaceans with primitively uniform thoracic segments (as also seen in branchiopods). Below we describe changes in *Ubx*–*abdA* expression in two groups of malacostracan crustaceans which carry different numbers of maxillipeds, peracarids and decapods.

In peracarids, the first, and sometimes second, of the eight thoracic segments bear limbs that have acquired several characteristics of feeding appendages. We find that the modification of these segments correlates with the repression of *Ubx*–*abdA* expression in those segments. In the mysid *Mysidium colombiae* (Mysida), the T1 limb has acquired the characteristic gnathal-like morphology of maxillipeds (Fig. 2a); a similar modification is seen in the most distal part of the T2 limb (Fig. 2a), although this limb still retains most morphological and functional characteristics of a swimming appendage. In comparison to leptostracans, the initial anterior



**Figure 1** Crustaceans with no maxillipeds. **a**, FP6.87 (brown) and engrailed (black) staining of *Artemia* before the morphological appearance of thoracic appendages. **b**, Higher-magnification view. *Ubx*–*abdA* expression begins in the first thoracic segment (T1) and extends into the more posterior regions. The anterior expression border is segmental, but will eventually extend to a slightly more anterior parasegmental border in the ventral neurogenic region<sup>7</sup>. **c**, Morphology of limbs in *Triops* larvae (3 days after hatching). Thoracic appendages (T1, T2) have similar morphology and are distinct from the appendages of gnathal segments (MxI is shown, MxII is reduced in size even further). **d**, FP6.87 staining

of *Triops* (2.5 days after hatching). **e**, Higher-magnification view, focused on cells at the limb bases. As in *Artemia*, *Ubx*–*abdA* expression begins at the T1 segment and extends posteriorly. **f**, Morphology of adult limbs in *Paranebalia*. All eight thoracic appendages have similar morphology; T1 is somewhat smaller, particularly its exopod (Ex). Gnathal appendages have very reduced endopods and exopods, but prominent endites (arrows). **g**, FP6.87 (brown) and engrailed (black) staining of *Paranebalia* as limbs are becoming visible. *Ubx*–*abdA* expression begins in T1 and extends posteriorly.

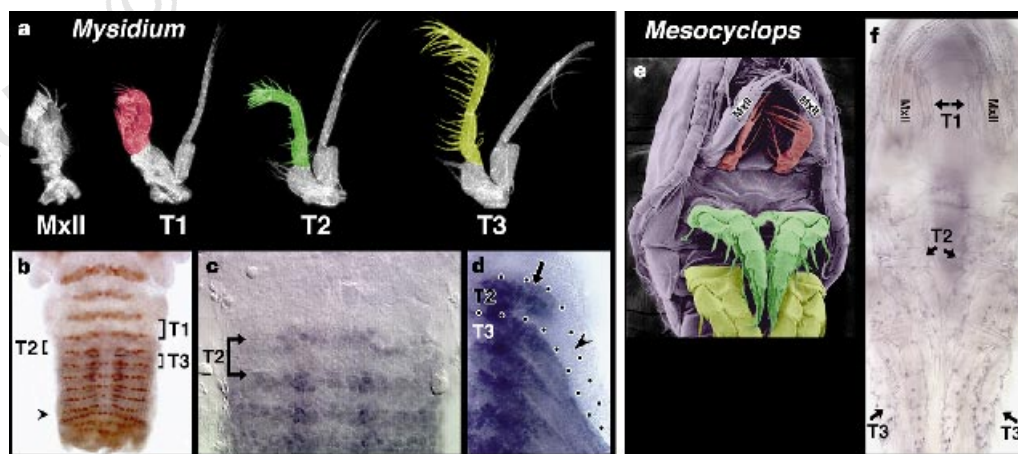
boundary of *Ubx-abdA* expression in mysids appears to be shifted backwards by an entire metameric unit. There is no FP6.87 staining in the T1 limb, weak staining in T2, and stronger staining in the more posterior regions of the trunk (Fig. 2b, c). We investigated this expression in more detail at different developmental stages. At all stages, we observed the same anterior borders of expression. At very early stages, when parasegments are only two cells wide, *Ubx-abdA* appear to be expressed or repressed uniformly within any given metameric unit, in both ectoderm and mesoderm (see posterior segments in Fig. 2b). Uniform early expression, however, becomes modulated within individual metameres during later development (differences in the level of expression, mosaic patterns; Fig. 2c, d). It is interesting to note that *Ubx-abdA* are expressed in the proximal portion of the T2 endopod, but excluded from its more distal part (Fig. 2d); it is this distal part of the T2 endopod that acquires gnathal-like characteristics (Fig. 2a). We have observed similar domains of *Ubx-abdA* expression in two other peracarids, a gammarid amphipod and an asellote isopod; in both cases the appearance of maxillipeds on T1 correlates with the repression of *Ubx-abdA* expression in that segment (data not shown).

Decapods are generally described as having three pairs of maxillipeds and five pairs of walking limbs in their thorax (hence the name Decapoda)<sup>9,12</sup>. Many of those limbs, however, show extensive structural variation, both between species and within the same species at different developmental stages. We examined two decapod genera showing different patterns of segmental specialization in the anterior thorax. The cleaner shrimps *Periclimenes yucatanicus* and *Periclimenes pedersoni* (Decapoda) show the typical decapod reduction and modification of the three most anterior pairs of thoracic limbs during embryonic development (Fig. 3a) and into adulthood (Fig. 3b). FP6.87 staining indicates that *Ubx-abdA* expression is excluded from the first three thoracic parasegments and limbs, is weak in T4, and stronger in more posterior segments (Fig. 3c, d). Thus, with respect to leptostracans, the anterior boundary of *Ubx-AbdA* distribution appears to have shifted backwards by three metameric units. In the

lobster *Homarus americanus* (Decapoda), the specialization of anterior thoracopods differs at embryonic and postembryonic stages. Adult lobsters have five pairs of large thoracic limbs (T4–T8) plus three pairs of reduced limbs (T1–T3; these are readily identifiable as maxillipeds). At hatching, however, only the T1 and T2 limbs appear to be distinctly reduced and similar to gnathal appendages. The T3 limb is morphologically similar to the more posterior walking legs and does not have the properties of a maxilliped (Fig. 3e). In *Homarus* embryos, FP6.87 staining is absent from the first two thoracic parasegments (including the T1 and T2 limbs), and strong in T3 and more posterior segments (Fig. 3f). Thus, the anterior boundary of embryonic expression of *Ubx-abdA* in *Homarus* appears to have shifted backwards by two metameric units relative to leptostracans, and corresponds to the morphological transition in thoracic limbs seen at hatching.

Maxillipeds are widely distributed among crustaceans and can also be found in non-malacostracan groups like copepods. We examined *Ubx-abdA* expression in two cyclopoid copepods, *Mesocyclops edax* and *Dioithona oculata* (Cyclopoida), both of which have a single pair of maxillipeds on T1 (Fig. 2e). Once again, this morphological arrangement appears to be associated with a posterior shift in *Ubx-AbdA* distribution: FP6.87 staining extends from T2 backwards, with no expression in the T1 limb, moderate expression in the T2 limb, and stronger expression further posterior (Fig. 2f). We observed a similar expression pattern in a diaptomid calanoid copepod (data not shown).

Collectively, our observations reveal a striking correlation between the expression of *Ubx-abdA* and the morphological specialization of corresponding segments. In particular, the repression of *Ubx* and *abdA* expression in anterior trunk segments is associated with the development of a gnathal-like identity in those segments. Bearing in mind the widely documented role of Hox genes in specifying segmental identity<sup>1,14,15</sup>, we suggest that this association may be direct and causal. A similar correlation between regional vertebral morphology and Hox expression has been reported among different species of vertebrates<sup>16,17</sup>; in that case,



**Figure 2** Crustaceans with one pair of maxillipeds. **a**, Morphology of adult limbs in *Mysidium*. MxII is second maxillary appendage; portions of the endopods of the first, second and third thoracic appendages (T1, T2 and T3) are shown in red, green and yellow respectively; thoracic exopods are the long right-most branches. The T1 limb is similar to that of gnathal appendages (as shown in MxII), with prominent endites and a reduction in particular elements of the endopod. The T2 limb has a 'mosaic' endopod, with a distal portion that resembles gnathal limbs and a proximal portion that resembles more posterior thoracic limbs. T3 to T8 limbs have similar morphology. **b**, FP6.87 (black) and engrailed (brown) staining during early *Mysidium* development. Staining is absent from T1, weak in T2, and strong in T3 and more posteriorly. Arrowhead marks T8, which at this stage is only 2 cells wide and shows uniform FP6.87

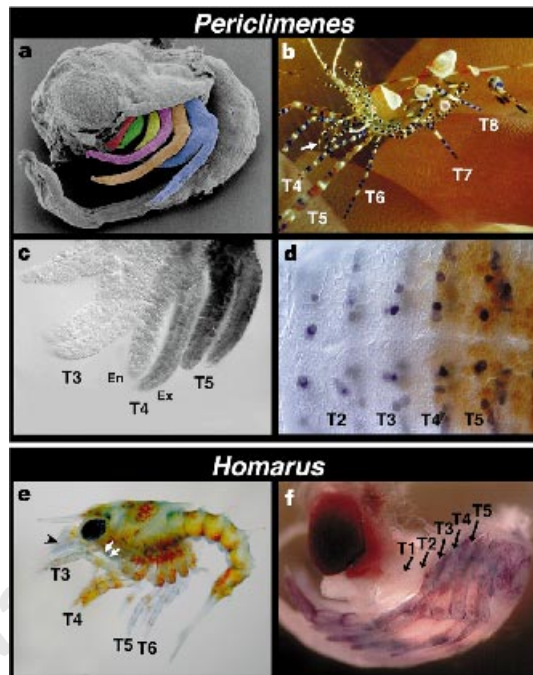
staining. **c**, FP6.87 staining in *Mysidium* at a slightly later time. Expression can be seen extending to the parasegment boundary in the more medial (neurogenic) region of posterior T1. At this stage FP6.87 staining shows modulated levels of expression within segments. **d**, FP6.87 staining in developing *Mysidium* limbs. In the T2 limb (outlined with dots), staining is seen in the proximal (black arrow) but not the distal (black arrowhead) portion of this limb; contrast this to the same portions of the T3 limb. **e**, Scanning electron micrograph of adult *Mesocyclops*. Limbs are colour-coded as in **a**. Maxillipeds on T1 are morphologically similar to gnathal appendages (MxII) and quite distinct from the remaining thoracic limbs (T2 shown). **f**, FP6.87 staining of larval stage of *Mesocyclops*. Nuclear staining appears punctate owing to the distant spacing of ectodermal cells in these mature copepod limbs.

however, the observed changes appear to be associated primarily with regional changes in vertebral numbers, not with the evolution of new vertebral identities.

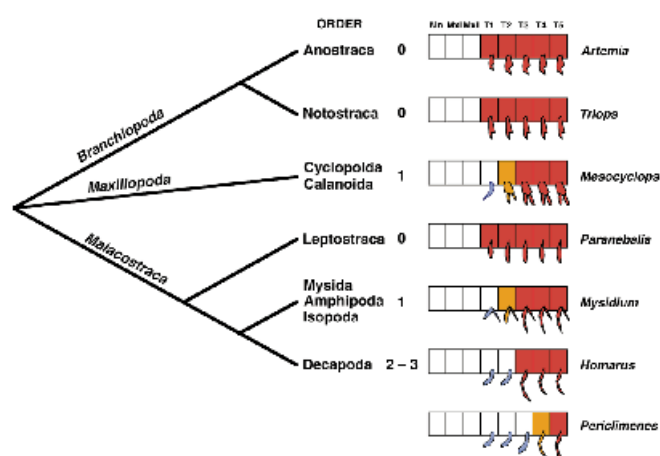
Maxilliped-bearing segments presumably acquire their distinct identity by expressing different combinations, patterns, or amounts of homeoproteins than do the other thoracic or gnathal segments. Studies in *Drosophila* have highlighted the importance of fine intrasegmental (mosaic) modulation in the expression patterns of Hox genes—it appears that different decisions on regional identity can be taken independently by different groups of cells within a segment<sup>15,18,19</sup>. The identity of a maxilliped-bearing segment could therefore be determined as a mosaic, with some parts of the segment retaining a thoracic identity and others becoming homeotically transformed to a gnathal fate. The spatially modulated distribution of *Ubx*–*AbdA* within the mysid T2 limb (Fig. 2d) and the resulting mosaic morphology of this appendage (Fig. 2a) are consistent with this hypothesis. In addition to spatial modulation, temporal changes in the expression of Hox genes are important because different decisions on regional identity can be taken at different developmental times<sup>18–20</sup>. Our results indicate that it is the early

patterns of *Ubx*–*abdA* expression (before and during the initial phases of limb growth) that correlate with the modification of anterior thoracic limbs into feeding appendages. Later changes in expression patterns are observed in *Artemia*<sup>7</sup> and *Paranebalia* (data not shown), where *Ubx*–*abdA* expression appears to retract from anterior thoracic segments as these segments mature. In both species this correlates with a reduction in the relative size of the corresponding thoracic limbs (Fig. 1f). Perhaps more interesting are cases like *Triops* and *Homarus*, where qualitative changes in the structure of anterior limbs can be seen during larval life. The late expression patterns of Hox genes in those species are not yet known.

The fossil record suggests that primitive crustaceans had a rather uniform series of thoracic segments, with no apparent specialization of anterior thoracic limbs<sup>21–23</sup>. Specialization of anterior thoracic appendages, however, is widespread among crustaceans today, and the phylogenetic distribution of these specializations suggests that similar morphological transformations have occurred independently several times during crustacean evolution (Fig. 4). Maxillipeds, for example, appear to have arisen independently in crustacean groups as diverse as malacostracans, copepods and



**Figure 3** Crustaceans with several pairs of maxillipeds. **a**, Scanning electron micrograph of late *Periclimenes yucatanicus* embryo showing the reduced size of the T1, T2 and T3 appendages (red, green and yellow respectively; T4 is pink, T5 is orange, T6 is blue, T7 and T8 are hidden more medially). **b**, Adult *P. yucatanicus* showing typical decapod morphology, with five pairs of walking legs (T4–T8) and three pairs of maxillipeds (T1–T3); arrow indicates the position of maxillipeds—only the T3 pair is clearly visible here. **c**, FP6.87 staining of developing *P. yucatanicus* limbs shows that staining is absent from T3 and more anterior limbs, weak in the T4 limb (En is endopod, Ex is exopod), and stronger in T5 and more posterior limbs. **d**, FP6.87 (brown) and engrailed (black) staining in developing neuromeres of *P. yucatanicus*; staining is weak in posterior T3 and anterior T4, and stronger more posteriorly. **e**, *Homarus* larvae 24 h after hatching. T3 is a well developed limb and is morphologically similar to more posterior walking legs. The maxillipeds of T1 and T2 are too small to be visible but their position is indicated by white arrowheads; grey arrowhead marks the antennal appendages. **f**, FP6.87 staining of *Homarus*; staining begins in T3 and extends posteriorly.



**Figure 4** Phylogenetic distribution of different patterns of segmental specialization and *Ubx*–*AbdA* expression among crustaceans. For each of the nine orders examined, we indicate the number of maxillipeds (0, 1, 2 or 3), and illustrate the different thoracic morphologies and segmental domains of *Ubx*–*AbdA* expression (dark orange for strong staining, light orange for weaker staining, and white for no staining) seen during early embryonic development. In each drawing three gnathal segments are depicted, followed by the first five thoracic segments. Maxillipeds are shown in blue, gnathal appendages are not shown. In each case, the early embryonic pattern of *Ubx*–*abdA* expression correlates with the eventual morphology of the thoracic segments; maxillipeds appear on those thoracic segments that do not express *Ubx*–*abdA*. The phylogenetic relationship of the three classes described here (Branchiopoda, Maxillopoda and Malacostraca) remains unresolved<sup>11,12</sup>. The ancestral crustacean body plan probably lacked maxillipeds, as did the ancestral malacostracan body plan<sup>21,21–23</sup>.



remipedes<sup>9,11</sup>. Our findings indicate that such convergent changes may have been achieved by similar developmental changes (involving similar posterior shifts in the expression boundary of *Ubx-abdA*) on several independent occasions. This suggests that, given a particular developmental system, there may be limited ways for achieving a particular morphological result.

It is often thought that the study of development mechanisms can provide a rigorous way to assess the homologous or convergent origin of morphological characteristics (see refs 7, 16, 24 for example). Our observations introduce one complication: this expectation will be true only in the extent to which convergent events are driven by distinguishable genetic changes. In this respect, it will be interesting to establish whether the observed convergent shifts in Hox expression can be attributed to distinguishable changes at a fine molecular-genetic level. □

## Methods

The monoclonal antibody against Ubx and AbdA (FP6.87) is described in ref. 13; the monoclonal antibodies against Engrailed (4D9 and 4F11) are described in ref. 25. Immunohistochemical staining is described in refs 25, 26. Cuticularized embryos and larvae were sonicated briefly to allow efficient penetration of reagents<sup>26</sup>. Further details on collection, rearing, fixation, sonication and staining are available upon request.

Received 25 March; accepted 10 June 1997.

- Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 567–570 (1978).
- Raff, R. A. & Kaufman, T. C. *Embryos, Genes, and Evolution* (Indiana University Press, Bloomington, 1991).
- Akam, M., Dawson, I. & Tear, G. Homeotic genes and the control of segment diversity. *Development* **104** (suppl.), 123–133 (1988).
- Bateson, W. *Materials for the Study of Variation, Treated with Especial Regard to Discontinuity in the Origin of Species* (Macmillan, London, 1894).
- Averof, M. & Akam, M. *HOM/Hox* genes of *Artemia*: implications for the origin of insect and crustacean body plans. *Curr. Biol.* **3**, 73–78 (1993).
- Averof, M., Dawes, R. & Ferrier, D. Diversification of arthropod *Hox* genes as a paradigm for the evolution of gene functions. *Sem. Dev. Cell Biol.* **7**, 539–551 (1996).
- Averof, M. & Akam, M. *Hox* genes and the diversification of insect and crustacean body plans. *Nature* **376**, 420–423 (1995).
- Panganiban, G., Sebring, A., Nagy, L. & Carroll, S. The development of crustacean limbs and the evolution of arthropods. *Science* **270**, 1363–1366 (1995).
- Calman, W. T. *Crustacea* (Black, London, 1909).
- Giesbrecht, W. in *Handwörterbuch der Naturwissenschaften* (ed. Dittler, R.) 800–840 (Fischer, Jena, 1931).
- Schram, F. R. *Crustacea* (Oxford University Press, Oxford, 1986).
- Brusca, R. C. & Brusca, G. J. *Invertebrates* 1–922 (Sinauer, Sunderland Massachusetts, 1990).
- Kelsh, R., Weinzierl, R. O. J., White, R. A. H. & Akam, M. Homeotic gene expression in the locust *Schistocerca*: an antibody that detects conserved epitopes in *Ultrabithorax* and *Abdominal-A* proteins. *Dev. Genet.* **15**, 19–31 (1994).
- McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
- Akam, M. *Hox* genes and the evolution of diverse body plans. *Phil. Trans. R. Soc. Lond. B* **349**, 313–319 (1995).
- Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebral axial morphology. *Development* **121**, 333–346 (1995).
- Gaunt, S. J. Conservation in the Hox code during morphological evolution. *Int. J. Dev. Biol.* **38**, 549–552 (1994).
- Castelli-Gair, J., Greig, S., Micklem, G. & Akam, M. Dissecting the temporal requirements for homeotic gene function. *Development* **120**, 1983–1995 (1994).
- Castelli-Gair, J. & Akam, M. How the Hox gene *Ultrabithorax* specifies two different segments: the significance of spatial and temporal regulation within metemeres. *Development* **121**, 2973–2982 (1995).
- Salser, S. J. & Kenyon, C. A. C. *elegans* Hox gene switches on, off, on, and off again to regulate proliferation, differentiation, and morphogenesis. *Development* **122**, 1651–1661 (1996).
- Briggs, D. E. G. The morphology, mode of life, and affinities of *Canadaspis perfecta* (Crustacea, Phyllocarida), Middle Cambrian, Burgess Shale, British Columbia. *Phil. Trans. R. Soc. Lond. B* **281**, 439–487 (1978).
- Muller, K. J. & Walossek, D. External morphology and larval development of the Upper Cambrian maxillopod *Bredocaris admirabilis*. *Foss. Strat.* **23**, 1–70 (1988).
- Walossek, D. The Upper Cambrian *Rehbachella* and the phylogeny of Branchiopoda and Crustacea. *Foss. Strat.* **32**, 1–202 (1993).
- Wagner, G. P. The biological homology concept. *Annu. Rev. Ecol. Syst.* **20**, 51–69 (1989).
- Patel, N. H., Kornberg, T. B. & Goodman, C. S. Expression of *engrailed* during segmentation in grasshopper and crayfish. *Development* **107**, 201–212 (1989).
- Patel, N. H. in *Methods in Cell Biology*, 44. *Drosophila Melanogaster: Practical Uses in Cell Biology* (eds Goldstein, L. S. B. & Fyrberg, E.) 445–487 (Academic, New York, 1994).

**Acknowledgements.** We thank M. Akam and F. Ferrari for discussion; K. Rützel and the Smithsonian Institution for hosting our collecting expeditions at Carrie Bow Cay in Belize; R. White for the FP6.87 antibody; G. Wyngaard for *Mesocyclops*; E. Chang for *Homarus* embryos; M. Sepanski for help with electron microscopy; A. Crittenden and L. Brown for technical assistance; and M. Palopoli, R. Chasan and S. Cohen for comments on the manuscript. This work was supported by the Wellcome Trust (M.A.), the Carnegie Institution of Washington, and the Howard Hughes Medical Institute (N.H.P.).

Correspondence and requests for materials should be addressed to N.H.P. (e-mail: npatel@midway.uchicago.edu).

## Recruitment of functional GABA<sub>A</sub> receptors to postsynaptic domains by insulin

Q. Wan\*, Z. G. Xiong†, H. Y. Man\*, C. A. Ackerley\*, J. Branton\*, W. Y. Lu†, L. E. Becker\*, J. F. MacDonald† & Y. T. Wang\*

\* Divisions of Pathology and Neuroscience, Hospital for Sick Children, and Departments of Pathology and † Physiology, University of Toronto, Toronto, Ontario M5G 1X8, Canada

Modification of synaptic strength in the mammalian central nervous system (CNS) occurs at both pre- and postsynaptic sites<sup>1,2</sup>. However, because postsynaptic receptors are likely to be saturated by released transmitter, an increase in the number of active postsynaptic receptors may be a more efficient way of strengthening synaptic efficacy<sup>3–7</sup>. But there has been no evidence for a rapid recruitment of neurotransmitter receptors to the postsynaptic membrane in the CNS. Here we report that insulin causes the type A  $\gamma$ -aminobutyric acid (GABA<sub>A</sub>) receptor, the principal receptor that mediates synaptic inhibition in the CNS<sup>8</sup>, to translocate rapidly from the intracellular compartment to the plasma membrane in transfected HEK 293 cells, and that this relocation requires the  $\beta$ 2 subunit of the GABA<sub>A</sub> receptor. In CNS neurons, insulin increases the expression of GABA<sub>A</sub> receptors on the postsynaptic and dendritic membranes. We found that insulin increases the number of functional postsynaptic GABA<sub>A</sub> receptors, thereby increasing the amplitude of the GABA<sub>A</sub>-receptor-mediated miniature inhibitory postsynaptic currents (mIPSCs) without altering their time course. These results provide evidence for a rapid recruitment of functional receptors to the postsynaptic plasma membrane, suggesting a fundamental mechanism for the generation of synaptic plasticity.

We used quantitative electron-microscopic analysis of immunogold labelling to investigate translocation of GABA<sub>A</sub> receptors in HEK 293 cells stably expressing the  $\alpha$ 1,  $\beta$ 2 and  $\gamma$ 2 subunits of rat GABA<sub>A</sub> receptors<sup>9</sup>, the most common subunit combination of native GABA<sub>A</sub> receptors in the mammalian CNS<sup>10</sup>. We found that, under control conditions, only a small number of immunogold-labelled  $\beta$ 2 subunits could be observed at the membrane surface, and that most labelling was localized intracellularly (Fig. 1a,c). Treatment with insulin dramatically increased the number of gold-labelled GABA<sub>A</sub> receptor subunits in the plasma membrane (Fig. 1b,c); the change could be detected as early as 10 min after insulin treatment (data not shown). An increase in receptor synthesis was not involved, as even 60 min of insulin treatment did not affect the total cellular expression of  $\beta$ 2 subunits (Fig. 1c). Pretreatment of cells with genistein, a membrane-permeable inhibitor of protein-tyrosine kinases<sup>11,12</sup>, prevented the insulin-induced translocation without altering basal GABA<sub>A</sub> receptor distribution (Fig. 1c), suggesting a requirement for activation of the insulin receptor tyrosine kinase<sup>13</sup>. To determine whether the insulin-induced receptor translocation leads to an increase in the number of functional GABA<sub>A</sub> receptors on the plasma-membrane surface, we recorded GABA<sub>A</sub> receptor-mediated whole-cell currents in these cells. Bath application of insulin (0.5  $\mu$ M) produced a significant increase in the amplitude of the currents, an effect not observed in cells pretreated with genistein (Fig. 1d). These results demonstrate a rapid increase in the number of functional GABA<sub>A</sub> receptors on the plasma-membrane surface as a result of receptor translocation by activation of insulin-receptor tyrosine kinase.

To determine which subunit(s) is required for insulin-induced

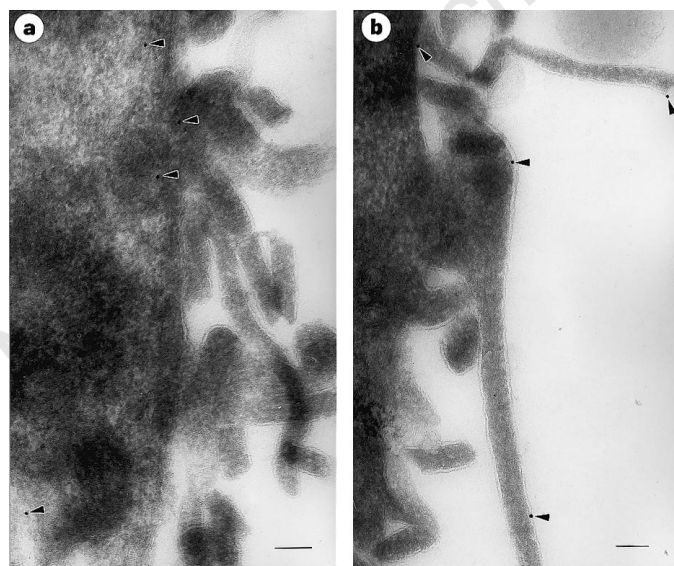


receptor translocation, HEK 293 cells were transiently transfected with rat GABA<sub>A</sub> receptor  $\beta 2$ ,  $\gamma 2$  and the epitope-tagged  $\alpha 1^{\text{FLAG}}$  subunit complementary DNAs in various combinations. Confocal scanning sections double-stained for surface and for entire cell staining demonstrated that a large proportion of  $\alpha 1\beta 2$  transfectants was surface-labelled (Fig. 2). However, among cells transfected with  $\alpha 1\gamma 2$ , or  $\alpha 1$  or  $\beta 2$  alone, few were surface-labelled. Insulin (0.5  $\mu\text{M}$ , 10 min) increased the proportion of surface-labelled cells in  $\alpha 1\beta 2$  and  $\beta 2$  by 29% and 103%, respectively, but not in  $\alpha 1\gamma 2$  or  $\alpha 1$  transfected cells (Fig. 2). Thus the  $\beta 2$  subunit seems to be required for the insulin-induced increase in receptor surface expression.

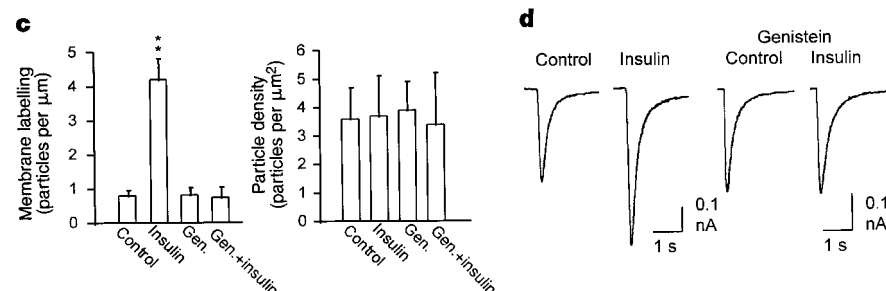
We next investigated whether insulin can effect the translocation of native CNS GABA<sub>A</sub> receptors *in situ* in cultured mouse hippocampal neurons. When native GABA<sub>A</sub> receptors were stained under permeabilized conditions with anti- $\beta 2/\beta 3$  antibody, basal levels of plasma membrane-associated receptors were observed in all cells. However, large numbers of the receptors were found to be intracellularly localized through the cell soma (Fig. 3A,a). Insulin (0.5  $\mu\text{M}$ , 10 min) increased the ratio of the receptor labelling on the plasma membrane to that in the cytoplasmic compartment (Fig. 3A, b and c), providing evidence for translocation of intracellularly localized receptors to the plasma membrane. Optical sections stained under non-permeabilized conditions demonstrated that insulin increased the labelling on the plasma-membrane surface (Fig. 3B), and that this increase in the surface labelling occurred preferentially on the proximate dendrites (Fig. 3B, c and d), sites where GABAergic synapses have been shown to be predominantly localized<sup>14</sup>. These results are consistent with a recruitment by insulin of GABA<sub>A</sub> receptors to the postsynaptic domains near the sites of presynaptic release. This possibility was further examined, using electron microscopy, in the CA1 neurons of control and insulin-

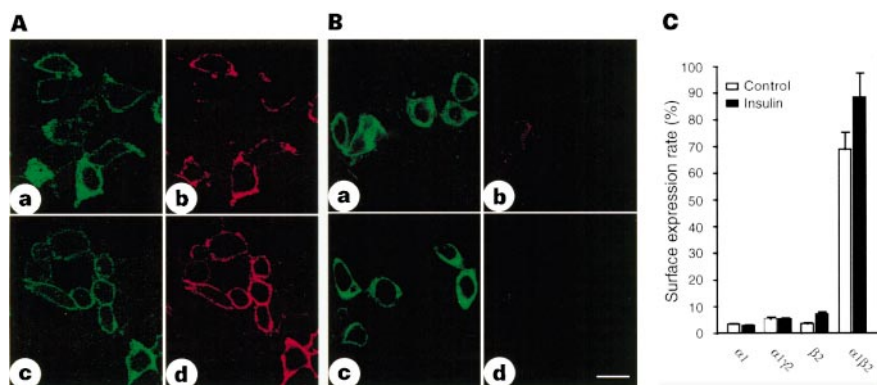
treated hippocampal slices. Under control conditions, most immunogold-labelled  $\beta 2/\beta 3$  subunits were found to be located along synaptic, dendritic and somatic membranes, with the highest concentration of receptors in the postsynaptic side of the synaptic junctions, consistent with the synaptic enrichment of the GABA<sub>A</sub> receptors previously reported<sup>15</sup>. Treatment of slices with insulin (0.5  $\mu\text{M}$ , 10 min) significantly increased the receptor density on both postsynaptic and dendritic membranes, but not on the somatic membrane (Fig. 3C and D, c). These results demonstrate that recruitment of GABA<sub>A</sub> receptors to the postsynaptic domains by insulin is not unique to neurons in primary culture, but can also occur in mature neurons *in situ*.

To determine whether the newly recruited surface receptors are functional, we examined effects of insulin on currents evoked by exogenously applied GABA in cultured neurons. Bath application of insulin (0.5  $\mu\text{M}$ ) potentiated the currents at both near- and sub-saturated GABA concentrations ( $26.4 \pm 5.3\%$  at 200  $\mu\text{M}$ ,  $15.6 \pm 7.9\%$  at 10  $\mu\text{M}$ ,  $n = 5$  neurons; Fig. 4A). Thus insulin increases the maximum response, providing support for an increase in the number of functional membrane GABA<sub>A</sub> receptors. To provide evidence for recruitment of functional GABA<sub>A</sub> receptors at postsynaptic junctions, we also investigated modulation by insulin of GABA<sub>A</sub> receptor-mediated mIPSCs. Within 10 min, insulin caused an increase of approximately 30% in the amplitude of the mIPSCs ( $-37 \pm 4.2$  control versus  $-47 \pm 2.9$  pA after insulin,  $P < 0.01$ ,  $n = 5$ ; Fig. 4B, a), but caused no change in either the rise time (10% to 90% of the peak,  $2.7 \pm 0.1$  versus  $2.6 \pm 0.2$  ms,  $P > 0.05$ ; Fig. 4B, b) or the decay time constant ( $\tau = 28.3 \pm 3$  versus  $26.8 \pm 2$  ms,  $P > 0.05$ ; Fig. 4B, b). Thus the results are consistent with an increased number of active postsynaptic receptors. Insulin was also found to increase mIPSC



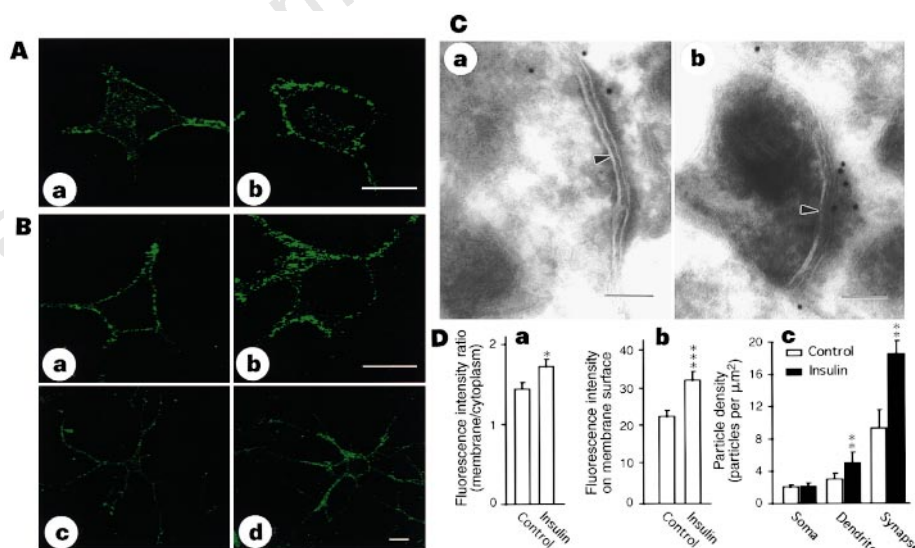
**Figure 1** Insulin induces a rapid translocation of GABA<sub>A</sub> receptors from intracellular compartments to the plasma membrane and potentiates GABA<sub>A</sub> receptor-mediated currents in HEK 293 cells stably transfected with cDNAs encoding rat GABA<sub>A</sub> receptor  $\alpha 1$ ,  $\beta 2$  and  $\gamma 2$  subunits. Electron micrographs of ultrathin cryosections labelled with immunogold using a monoclonal antibody against the rat  $\beta 2/\beta 3$  GABA<sub>A</sub> receptor subunits show subcellular distribution of the  $\beta 2$  subunits in control (a) and after a 60-min insulin (0.5  $\mu\text{M}$ ) stimulation (b). Immunogold-labelled receptor proteins are indicated by arrowheads; scale bars, 0.1  $\mu\text{m}$ . c, Quantification of membrane and total cellular expression of  $\beta 2$  subunits. Cells were incubated for 60 min in the extracellular solution (control) or in solution supplemented with either 0.5  $\mu\text{M}$  insulin or 50  $\mu\text{M}$  genistein (Gen.). For genistein pretreatment (Gen. + insulin), cells were first treated with 50  $\mu\text{M}$  genistein for 10 min, and then incubated for 60 min in the genistein solution supplemented with 0.5  $\mu\text{M}$  insulin. Values are means  $\pm$  s.e. from 3 experiments (25 cells per experiment) and were analysed using the unpaired Student *t*-test (\*\* denotes  $P < 0.01$ ). d, Recordings of representative individual GABA<sub>A</sub> receptor-mediated whole-cell currents induced by puffing GABA (100  $\mu\text{M}$ , 100 ms) onto the transfected cells at a holding membrane potential of  $-60$  mV before (control) and 10 min after addition of 0.5  $\mu\text{M}$  insulin. For genistein treatment, cells were incubated in 50  $\mu\text{M}$  genistein before the start of whole-cell recording and exposed to the drug throughout recording period. Insulin increased the amplitudes of the currents by 40% ( $-376.1 \pm 106.6$  pA for control versus  $-526.3 \pm 145.3$  pA 10 min after insulin,  $P < 0.05$ ,  $n = 8$ ), and this effect was prevented by genistein pretreatment ( $-263 \pm 83.7$  versus  $-261 \pm 81.6$  pA;  $P > 0.05$ ,  $n = 6$ ).





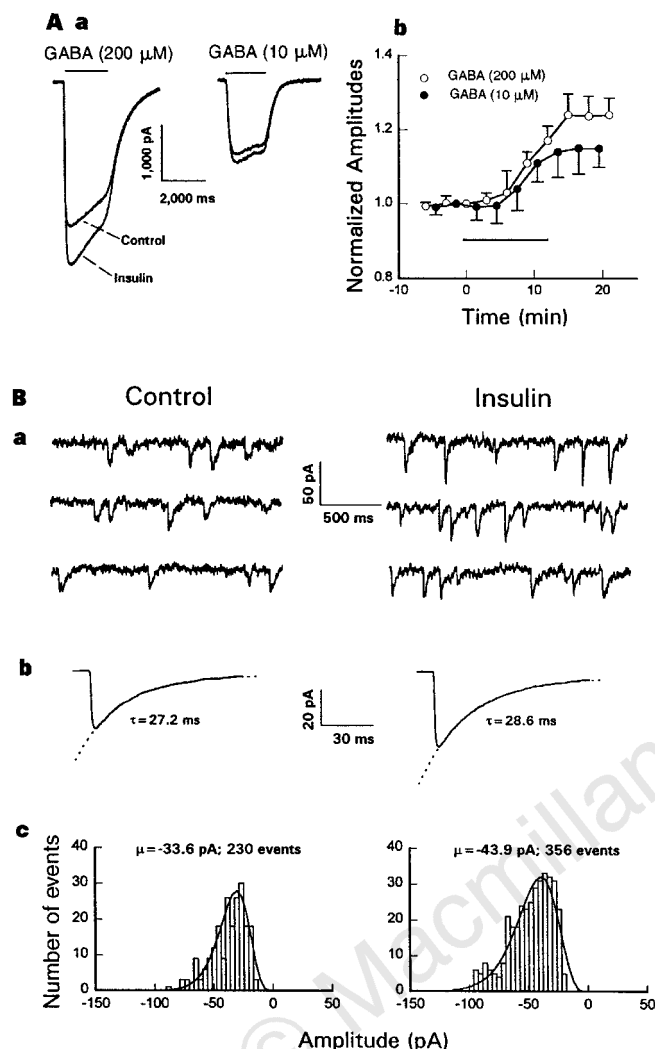
**Figure 2** The β-subunit is required for insulin-mediated translocation of recombinant GABA<sub>A</sub> receptors transiently expressed in HEK 293 cells. **A, B**, Confocal optical sections (1 μm) of cells transfected with α1<sup>FLAG</sup>β2 (**A**) or α1<sup>FLAG</sup>γ2 (**B**) before (**a** and **b**) and after (**c** and **d**) 10-min insulin treatment (0.5 μM). Double-immunofluorescence staining was done by first staining (under non-permeabilized conditions) for the FLAG epitope tag inserted in the extracellular domain (near the N terminus) of α1 (**b** and **d**, red) followed by permeabilized staining for the main intracellular loop of the same subunit (**a** and **c**, green). **C**, Effects of insulin on cell surface expression. For cells transfected with α1<sup>FLAG</sup>,

α1<sup>FLAG</sup>γ2 or α1<sup>FLAG</sup>β2 subunit combinations, the rate of surface expression was calculated as the number of surface-labelled cells (labelled with anti-Flag antibody) as a percentage of total transfected cells (labelled with polyclonal anti-α antibody). For cells transfected with the β-subunit alone, surface-labelled cells and transfected cells were identified by staining with antibody to β2/β3 subunits under non-permeabilized and permeabilized conditions, respectively. Each combination was repeated in 3 experiments, and 500–1,000 transfected cells were counted in each combination in each experiment.



**Figure 3** Insulin causes membrane translocation and clustering of native GABA<sub>A</sub> receptors in CNS neurons. Control (**a**) and insulin-treated (**b**; 0.5 μM, 10 min) cultured mouse hippocampal neurons (**A–C**) were stained with the β2/β3 monoclonal antibody. **A**, Optical sections (1 μm) of the cell soma stained under non-permeabilized conditions show that insulin induces a relocation of intracellular GABA<sub>A</sub> receptors to the plasma membrane. **B**, Optical sections scanned through the cell soma show that, under non-permeable conditions, GABA<sub>A</sub> receptors expressed on the membrane surface are selectively labelled and insulin increases the surface labelling (**a** and **b**). The same neurons shown in **a** and **b** are below in stacked serial optic sections of the entire cell body. Under basal conditions (**c**), numerous small clusters of GABA<sub>A</sub> receptors are distributed diffusely over the entire cell, including remote processes. After insulin treatment (**d**), receptors are concentrated in large clusters on the proximate dendrites. Scale bars, 10 μm for **A** and **B**. **C**, Insulin increases synaptic GABA<sub>A</sub> receptor density of

mature CNS neurons *in situ*. Electron micrographs show immunogold-labelled GABA<sub>A</sub> receptor β2/β3 subunits in the postsynaptic junctions (arrowheads) in the hippocampal CA1 region from a control slice (**a**) and a slice treated with 0.5 μM insulin for 10 min at 35 °C (**b**). Scale bars, 0.1 μm. **D**, Effect of insulin on surface expression of native GABA<sub>A</sub> receptors. **a**, Quantification of the ratio of fluorescence intensity of β2/β3 subunit staining on the membrane to that in cytoplasm from cultured neurons stained under permeabilized conditions (*n* = 25 neurons). **b**, Quantification of fluorescence intensity on the membrane surface from cultured neurons stained under non-permeabilized conditions (*n* = 45 neurons). **c**, Quantification of immunoparticle densities associated with the β2/β3 subunits of native GABA<sub>A</sub> receptors in mature hippocampal CA1 neurons *in situ*, categorized by membrane region; *n* = 25, 25 and 100 for somatic, dendritic and synaptic membranes, respectively. \**P* < 0.05, \*\* 0.01 and \*\*\* 0.001.



frequency (Fig. 4B, c); although this could possibly be due to a higher rate of quantal release<sup>16,17</sup>, it could also be fully accounted for by the recruitment of mIPSCs, previously below the threshold of detection, owing to an increase in the number of active receptors<sup>18</sup>. Finally, in CA1 hippocampal neurons in brain slices prepared from adult rats, insulin enhanced the amplitude of the mIPSCs by 23% ( $20.1 \pm 1.9$  versus  $24.6 \pm 2.8$  pA,  $P < 0.01$ ,  $n = 4$ ) but not their time course ( $\tau = 21.2 \pm 2.6$  versus  $22.6 \pm 4.1$  ms,  $P > 0.05$ ). These results suggest that the insulin-induced increase in the number of functional postsynaptic GABA<sub>A</sub> receptors occurs in mature neurons *in situ*, indicating that the receptor relocation is likely to be of physiological importance.

Our results have demonstrated that insulin can regulate the surface expression of GABA<sub>A</sub> receptors and thereby modulate GABA<sub>A</sub>-receptor-mediated synaptic inhibition. Insulin and insulin receptors are expressed at high levels in discrete regions within the CNS, and insulin is released by depolarization in cultured CNS neurons<sup>13</sup>. Insulin has been shown to act as a neuromodulator of many brain functions, such as food intake<sup>13</sup>. The regulation of food intake by brain insulin may be mediated in part through GABA<sub>A</sub> receptors, as insulin-induced hyperphagia in freely moving rats can be blocked by antagonists of GABA<sub>A</sub> receptors applied in the ventromedial hypothalamic region<sup>19</sup>. The translocation of GABA<sub>A</sub> receptors by insulin may also be important outside the CNS. In pancreatic islet, GABA is expressed at high levels in insulin-secreting  $\beta$ -cells<sup>20</sup>, and functional GABA<sub>A</sub> receptors are present in glucagon-secreting  $\alpha$ -cells<sup>21</sup>. It has been reported that glucose inhibition of glucagon secretion is mediated through the activation of GABA<sub>A</sub>

**Figure 4** Potentiation of GABA<sub>A</sub> receptor-mediated current responses by insulin in cultured hippocampal neurons. **A**, Insulin potentiates whole-cell GABA currents in perforated patch configuration at a holding membrane potential of  $-60$  mV. Currents were induced in the same neuron by alternate perfusion of GABA at saturated and non-saturated concentrations through a multi barrel fast perfusion system. The doses were determined from full dose-response curves constructed for three neurons under control conditions; 10 and 200  $\mu$ M are doses producing approximately half and maximum responses, respectively (data not shown). **a**, Superimposed sample current traces before and 10 min after 0.5  $\mu$ M insulin application. **b**, Normalized peak currents ( $I_{\text{insulin}}/I_{\text{control}}$ ) from 5 individual neurons. Insulin was applied in the bath during the period indicated by the bar. **B**, Whole-cell recordings of GABA<sub>A</sub> receptor-mediated mIPSCs at a holding potential of  $-60$  mV before (control) and 8 min after addition of insulin (0.5  $\mu$ M). **a**, Examples of consecutive traces of mIPSCs. **b**, Averages of 100 individual mIPSCs. The decay of mIPSCs before and after insulin were well fitted by a single exponential. **c**, Histograms showing amplitude distributions of a 3-min recording of mIPSCs. The histograms were fitted using an equation described previously<sup>30</sup>. Bath application of insulin for 5–10 min increased the amplitude of mIPSCs without altering their time course. Note that insulin also increased the frequency of mIPSCs ( $0.94 \pm 0.17$  versus  $1.6 \pm 0.25$  events  $s^{-1}$ ,  $n = 5$ ,  $P < 0.01$ )

receptors on the  $\alpha$ -cells, and this does not seem to be associated with any detectable increase in GABA release<sup>21</sup>. Because secretion of insulin in the islet increases in the presence of high concentrations of glucose, our results suggest that insulin may mediate glucose feedback regulation of glucagon release by increasing the activity of GABA<sub>A</sub> receptors in  $\alpha$ -cells. Thus the observed modulation of GABA<sub>A</sub> receptors by insulin may be important for the regulation of food intake and glucose homeostasis under physiological conditions, as well as in diseases such as diabetes and obesity.

Receptor translocation has important implications for synaptic function. Evidence from both experimental and modelling studies supports the view that GABA released from a single vesicle nearly saturates postsynaptic GABA<sub>A</sub> receptors at CNS synapses<sup>3,6,7</sup>. Therefore, an increase in quantal transmitter release would have little effect on the amplitude of the inhibitory synaptic potential. In contrast, the insertion of additional functional GABA<sub>A</sub> receptors into the postsynaptic membrane would serve as an effective means of enhancing inhibitory synaptic transmission. Thus the rapid insulin-induced translocation of GABA<sub>A</sub> receptors to the postsynaptic domain of inhibitory synapses provides a mechanism for the generation of synaptic plasticity in these synapses<sup>22</sup>, and may also have implications for the proposed recruitment of silent synapses following the long-term potentiation of glutamate-receptor-mediated synaptic transmission<sup>18,23–25</sup>. □

#### Methods

**Immunogold staining and electron-microscopic examination.** HEK cells stably transfected with cDNAs encoding the  $\alpha 1$ ,  $\beta 2$  and  $\gamma 2$  subunits of rat

GABA<sub>A</sub> receptors (4D4 cells)<sup>9</sup> were provided by D. Carter. Transcription of all three subunit mRNAs in the cell line was confirmed by reverse transcription–polymerase chain reaction (RT–PCR) (data not shown). Cells of passages 14–19 were plated in 100-mm culture dishes and used at 80% confluence. Cells were rinsed in PBS and incubated in the extracellular solution (control) or in the extracellular solution supplemented with 0.5 μM insulin for 10–60 min. Vibratome hippocampal slices (300 μm) were prepared from adult Sprague–Dawley rats. After a recovery period of 1 h artificial cerebrospinal fluid (ACSF)<sup>26</sup> at 35 °C, slices were incubated in ACSF containing 0.5 μM insulin for 10 min. After treatment, cells or slices were immediately fixed in 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.2) for 1 h. Pelleted cells were then embedded in gelatin. Embedded cells or dissected CA1 region of the slices were infused with 2.3 M sucrose for several hours and frozen in liquid nitrogen. Sections (100 nm) were then cut and immunogold-labelled using the monoclonal antibody bd-17 against rat β2/β3 subunits and 10-nm gold particles conjugated to goat anti-mouse IgG. In some cases, primary antibody was omitted, or replaced with a monoclonal antibody (bd-24) that recognizes human, but not rat, GABA<sub>A</sub>-receptor α-subunit<sup>27</sup>, to show the specificity of β2/β3 labelling. Non-transfected 293 cells were also stained with anti-β2/β3 antibody as a negative control. Images were captured with a CCD camera in the transmission electron microscope, and gold particles were quantified using the morphometric program NIH Image 1.59.

**Cell transfection and neuronal cultures.** The rat GABA<sub>A</sub>-receptor subunit cDNAs, pcDNA1-β2 and the pcDNA3-γ2 (short form of γ2), were provided by C. Kaufman and D. Gunnersen (Laboratory of Neuroscience, NIDDK). The Flag-tagged rat α1 subunit cDNA (pcDNA3-α1<sup>FLAG</sup>) was provided by J. H. Steinbach. This was generated by inserting the Flag sequence (KDYKDDDDKL) in the extracellular N-terminal domain between amino acids 33 and 34 of the translated sequence. This epitope tag did not alter either expression or function of the recombinant GABA<sub>A</sub> receptor, and the tagged α1 subunit protein was recognized on the surface of intact cells using the M2 monoclonal anti-FLAG antibody<sup>28</sup>. HEK 293 cells were plated onto collagen-coated 22-mm glass coverslips set in standard 35-mm culture dishes and transfected with 2 μg of each GABA<sub>A</sub>-receptor subunit plasmid using Lipofectamine (GIBCO) according to the manufacturer's protocol. Cultured mouse hippocampal neurons were grown as previously described<sup>29</sup>.

**Immunofluorescent confocal microscopy.** HEK 293 cells at 48 h post-transfection or neurons following 8–20 days in culture were fixed with 4% paraformaldehyde in PBS for 10 min and permeabilized with 0.25% Triton X-100 in PBS for 10 min, or fixed without permeabilization. GABA<sub>A</sub> receptors were stained with bd-17 monoclonal antibody recognizing rat β2/β3 subunits (15 μg ml<sup>-1</sup>) and with FITC-conjugated secondary antibody. For double-labelling experiments, cells were first stained with monoclonal anti-FLAG epitope (16 μg ml<sup>-1</sup>) and rhodamine-conjugated anti-mouse antibody without permeabilization. After a further 10 min fixation and 10 min permeabilization, cells were stained with a polyclonal antibody recognizing the rat α-subunit (1:200) and with FITC-conjugated anti-rabbit antibody. Specific staining was tested by either replacing the monoclonal primary antibody with an unrelated antibody, the monoclonal anti-α1 subunit of the human GABA<sub>A</sub> receptor (clone bd-24)<sup>27</sup>, or by omitting the primary antibody in the case of the polyclonal antibody. Subcellular localization of immunofluorescently labelled receptors was examined with a Leica TCS-4D confocal microscope, and fluorescence intensity was quantified using the Scion ImagePC software.

**Electrophysiology.** Whole-cell recordings of stably transfected HEK 293 cells (4D4 cells) and cultured neurons were made using an Axopatch-1B or -1D amplifier (Axon Instruments, Foster City, CA) at room temperature (20–22 °C). Culture medium was exchanged completely with standard extracellular solution 1–2 h before recording. The extracellular solution contained (in mM): 140 NaCl, 1.3 CaCl<sub>2</sub>, 5.4 KCl, 1 MgCl<sub>2</sub>, 25 HEPES, 33 glucose, 0.0005 TTX (pH 7.4 using NaOH, 320–335 mosm). The intracellular solution contained (in mM): 120 CsCl, 35 CsOH, 2 MgCl<sub>2</sub>, 1 CaCl<sub>2</sub>, 11 EGTA, 2 TEA, 10 HEPES, 4 ATP, pH 7.3. For perforated-patch recording, 100 μg ml<sup>-1</sup> nystatin (Sigma) was added to the intracellular solution. Standard whole-cell recordings of CA1 neurons were performed in adult rat hippocampal slices perfused with ACSF at 35 °C in a submerged recording chamber. For recording of mIPSCs, 2 mM Mg<sup>2+</sup> (plus AP5 for neurons in slices) and 20 μM CNQX were added to the

extracellular solution to inhibit NMDA and AMPA receptor-mediated excitatory synaptic currents. The mIPSCs were abolished by the addition of 10 μM GABA<sub>A</sub> receptor antagonist bicuculline. mIPSCs were first recorded on videotape and later acquired and analysed off-line using the SCAN program (Strathclyde Software).

**Drugs and materials.** Insulin (purified pancreatic bovine, Zn<sup>2+</sup>-free) was provided by C. C. Yip (Banting & Best Institute, Toronto, Canada). Monoclonal antibody bd-17, which recognizes extracellular epitopes on both β2 and β3 subunits of the rat GABA<sub>A</sub> receptor<sup>27</sup>, was purchased from Boehringer Mannheim Biochem (Canada). The polyclonal antibody against the main intracellular loop between transmembrane domains III and IV of rat α1 subunit of the GABA<sub>A</sub> receptor was purchased from PharMingen (San Diego). The monoclonal antibody to the Flag epitope (M2) was from Eastman Kodak Scientific Imaging System (New Haven).

Received 27 February; accepted 12 May 1997.

- Jessell, T. M. & Kandel, E. R. Synaptic transmission: A bidirectional and self-modifiable form of cell–cell communication. *Neuron* **10** (suppl.), 1–10 (1993).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Busch, C. & Sakmann, B. Synaptic transmission in hippocampal neurons: numerical reconstruction of quantal IPSCs. *Cold Spring Harb. Quant. Biol.* **55**, 69–80 (1990).
- Redman, S. Quantal analysis of synaptic potentials in neurons of the central nervous system. *Physiol. Rev.* **70**, 165–198 (1990).
- Tong, G. & Jahr, C. E. Multivesicular release from excitatory synapses of cultured hippocampal neurons. *Neuron* **12**, 51–59 (1994).
- Edwards, F. A. Anatomy and electrophysiology of fast central synapses lead to a structural model for long-term potentiation. *Physiol. Rev.* **75**, 759–787 (1995).
- Mody, I., Koninck, Y., Otis, T. S. & Soltesz, I. Bridging the cleft at GABA synapses in the brain. *Trends Neurosci.* **17**, 517–525 (1994).
- Macdonald, R. L. & Olsen, R. W. GABA<sub>A</sub> receptor channels. *Annu. Rev. Neurosci.* **17**, 569–602 (1994).
- Hamilton, B. J. et al. Stable expression of cloned rat GABA<sub>A</sub> receptor subunits in a human kidney cell line. *Neurosci. Lett.* **153**, 206–209 (1993).
- McKernan, R. M. & Whiting, P. J. Which GABA<sub>A</sub>-receptor subtypes really occur in the brain? *Trends Neurosci.* **19**, 139–143 (1996).
- O'Dell, T. J., Kandel, E. R. & Grant, S. G. Long-term potentiation in the hippocampus is blocked by tyrosine kinase inhibitors. *Nature* **353**, 558–560 (1991).
- Wang, Y. T. & Salter, M. W. Regulation of NMDA receptors by protein-tyrosine kinases and phosphatases. *Nature* **369**, 233–235 (1994).
- Wozniak, M., Rydzewski, B., Baker, S. P. & Raizada, M. K. The cellular and physiological actions of insulin in the central nervous system. *Neurochem. Int.* **22**, 1–10 (1993).
- Benson, D. L. & Cohen, P. A. Activity-independent segregation of excitatory and inhibitory synaptic terminals in cultured hippocampal neurons. *J. Neurosci.* **16**, 6424–6432 (1996).
- Nusser, Z., Roberts, J. D., Baude, A., Richards, J. G. & Somogyi, P. Relative densities of synaptic and extrasynaptic GABA<sub>A</sub> receptors on cerebellar granule cells as determined by a quantitative immunogold method. *J. Neurosci.* **15**, 2948–2960 (1995).
- Thompson, S. M., Copogna, M. & Scanziani, M. Presynaptic inhibition in the hippocampus. *Trends Neurosci.* **16**, 222–227 (1993).
- Cohen, G. A., Doze, V. A. & Madison, D. V. Opioid inhibition of GABA release from presynaptic terminals of rat hippocampal interneurons. *Neuron* **9**, 325–335 (1992).
- Liao, D., Hessler, N. A. & Malinow, R. Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. *Nature* **375**, 400–404 (1995).
- Kamatchi, G. L., Bhakthavatsalam, P., Chandra, D. & Bapna, J. S. Inhibition of insulin hyperphagia by gamma aminobutyric acid antagonists in rats. *Life Sci.* **34**, 2297–2301 (1984).
- Garry, D. J., Sorenson, R. L., Elde, R. P., Maley, B. E. & Madsen, A. Immunohistochemical colocalization of GABA and insulin in beta-cells of rat islet. *Diabetes* **35**, 1090–1095 (1986).
- Rorsman, P. et al. Glucose-inhibition of glucagon secretion involves activation of GABA<sub>A</sub>-receptor chloride channels. *Nature* **341**, 233–236 (1989).
- Otis, T. S., De Koninck, Y. & Mody, I. Lasting potentiation of inhibition is associated with an increased number of gamma-aminobutyric acid type A receptors activated during miniature inhibitory postsynaptic currents. *Proc. Natl Acad. Sci. USA* **91**, 7698–7702 (1994).
- Kullmann, D. M. Amplitude fluctuations of dual-component EPSCs in hippocampal pyramidal cells: Implications for long-term potentiation. *Neuron* **12**, 1111–1120 (1994).
- Edwards, F. LTP is a long term problem. *Nature* **350**, 271–272 (1991).
- Isaac, J. T., Nicoll, R. A. & Malenka, R. C. Evidence for silent synapses: implications for the expression of LTP. *Neuron* **15**, 427–434 (1995).
- Wang, Y. T., Neuman, R. S. & Bieger, D. Nicotinic cholinergic-mediated excitation in ambigal motoneurons of the rat. *Neuroscience* **40**, 759–767 (1991).
- Ewert, M., Shivers, B. D., Luddens, H., Mohler, H. & Seeburg, P. H. Subunit selectivity and epitope characterization of mAbs directed against the GABA<sub>A</sub>/benzodiazepine receptor. *J. Cell Biol.* **110**, 2043–2048 (1990).
- Ueno, S., Zorumski, C., Bracamontes, J. & Steinbach, J. H. Endogenous subunits can cause ambiguities in the pharmacology of exogenous gamma-aminobutyric acid A receptors expressed in human embryonic kidney 293 cells. *Mol. Pharmacol.* **50**, 931–938 (1996).
- MacDonald, J. F., Mody, I. & Salter, M. W. Regulation of N-methyl-D-aspartate receptors revealed by intracellular dialysis of murine neurons in culture. *J. Physiol. (Lond.)* **414**, 17–34 (1989).
- Bekkers, J. M., Richerson, G. B. & Stevens, C. F. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc. Natl Acad. Sci. USA* **87**, 5359–5362 (1990).

**Acknowledgements.** We thank M. Wojtowicz and F. Dorri for assistance with slice experiments; C. Kaufman and D. Gunnersen for GABA<sub>A</sub> receptor β2 and γ2 subunit cDNAs; J. H. Steinbach for the α1<sup>FLAG</sup> cDNA; D. Carter for the 4D4 cells; C. C. Yip for insulin; and M. W. Salter for helpful comments on the manuscript. This work was supported by the Medical Research Council of Canada and Neuroscience Network of Centers of Excellence of Canada. Y.T.W. is a research scholar of the Heart and Stroke Foundation of Canada/Ontario; Z.G.X. is a MRC-NCE fellow and W.Y.L. is a NCE fellow. H.Y.M. is supported by a studentship from the Canadian Foundation for the Study of Infant Death.

Correspondence and requests for materials should be addressed to Y.T.W. (e-mail: y.t.wang@utoronto.ca).



# Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module

Deborah Fass<sup>\*†</sup>, Stephen Blacklow<sup>\*†‡</sup>, Peter S. Kim<sup>\*†</sup> & James M. Berger<sup>†</sup>

<sup>\*</sup>Howard Hughes Medical Institute, Department of Biology,

Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

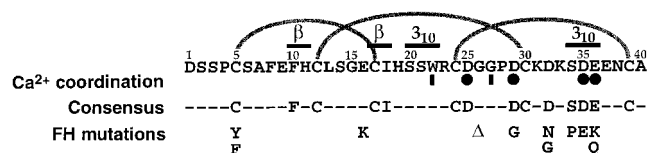
<sup>†</sup>Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

<sup>‡</sup>Present address: Department of Pathology, Stanford University, 300 Pasteur Drive, Stanford, California 94305-5324, USA.

The low-density lipoprotein receptor (LDLR) is responsible for the uptake of cholesterol-containing lipoprotein particles into cells<sup>1,2</sup>. The amino-terminal region of LDLR, which consists of seven tandemly repeated, ~40-amino-acid, cysteine-rich modules (LDL-A modules), mediates binding to lipoproteins<sup>3,4</sup>. LDL-A modules are biologically ubiquitous domains, found in over 100 proteins in the sequence database<sup>5</sup>. The structure of ligand-binding repeat 5 (LR5) of the LDLR, determined to 1.7 Å resolution by X-ray crystallography and presented here, contains a calcium ion coordinated by acidic residues that lie at the carboxy-terminal end of the domain and are conserved among LDL-A modules. Naturally occurring point mutations found in patients with the disease familial hypercholesterolaemia<sup>6</sup> alter residues that directly coordinate Ca<sup>2+</sup> or that serve as scaffolding residues of LR5.

Familial hypercholesterolaemia (FH) is a genetic disease caused by defects in LDLR function<sup>6</sup>. Normally the LDL-A modules of the LDLR mediate binding to the lipoproteins LDL and β-VLDL<sup>3,4</sup>, which are then imported into the cell by receptor-mediated endocytosis<sup>1,2</sup>. Mutations of the LDLR that affect this process result in failure to clear lipoproteins from the circulation, pathologically elevated blood cholesterol and premature heart disease<sup>6</sup>. Many point mutations that cause FH map to the fifth LDL-A module of the LDLR (LR5), particularly to a cluster of acidic residues near the carboxy-terminal end of the module (Fig. 1).

To investigate the molecular basis of FH, we determined the structure of LR5 by X-ray crystallography (Fig. 2). Crystals of LR5 were grown from ammonium sulphate with sucrose as a cryoprotectant (see Methods). Reflection data were collected at -180 °C and



**Figure 1** Sequence of LR5. Numbering refers to residue position in the LR5 peptide<sup>9</sup> and corresponds to that used in the text. Above the sequence, disulphide connectivity is indicated by loops, and locations of secondary structure elements are shown by black bars. Below the sequence, black circles highlight residues with side chains that participate in Ca<sup>2+</sup> coordination. Vertical bars indicate residues with backbone carbonyl oxygens that contribute to Ca<sup>2+</sup> coordination. Residues explicitly shown in the consensus sequence are conserved in at least 6 of the 7 amino-terminal LDLR modules<sup>24</sup>. The positions and identities of mutations causing FH<sup>6</sup> are displayed under the consensus sequence. The triangle indicates deletion of a residue.

phases were determined by multiple isomorphous replacement (MIR) with heavy-atom compounds (Table 1). The working and free *R*-factors to 1.7 Å resolution are 20.9 and 24.0%, respectively, with good protein geometry (Table 1).

The striking feature of the LR5 structure is that it is organized around a calcium ion. Although the overall topology of LR5 is similar to that of homologous modules 1 and 2 determined by NMR<sup>7,8</sup>, the NMR structures do not show evidence of Ca<sup>2+</sup> binding. The residues in LR5 that coordinate the Ca<sup>2+</sup> ion through their side chains are Asp 25, Asp 29, Asp 35 and Glu 36 (Figs 1, 3), the same acidic residues that are conserved among the LDL-A modules. The backbone carbonyls of Trp 22 and Gly 27 complete the octahedral, or six-ligand, coordination geometry (Fig. 3). The ion binding site explains the Ca<sup>2+</sup> requirement for proper folding and disulphide bond formation of LR5 (ref. 9) and provides a rationale for the Ca<sup>2+</sup> requirement in lipoprotein binding<sup>10</sup>.

The effects of FH mutations that occur in module 5 of the LDLR (Fig. 1) can be explained by the LR5 structure. These mutations fall into two categories: elimination or misplacement of direct Ca<sup>2+</sup> ligands, and removal of hydrogen bonds or disulphide bonds that help maintain the backbone topology. In the first category, mutation of Asp 29, Asp 35 or Glu 36 destroys the Ca<sup>2+</sup>-binding site. Deletion of Gly 26 alters the spacing between the Ca<sup>2+</sup> coordinating residues. In the second category, Glu 16 is in a position to form a hydrogen bond with the backbone amides of residues 16 and 31 by using both of its carboxylate oxygens, as well as with the Ser 14 side-chain hydroxyl group. This hydrogen-bonding network of Glu 16 is apparently crucial for stabilizing the tertiary structure of the module. Similarly, the Asp 32 side chain makes a hydrogen bond

**Table 1** Structure determination and refinement

| Derivative                       | Resolution (Å) | Completeness (%)                                      | $R_{\text{sym}}^*$ | $R_{\text{iso}}^\dagger$ | No. of sites                    | Phasing power‡ | Cullis $R_s^\S$                    | Cullis $R_{\text{anomalous}}^\S$ |
|----------------------------------|----------------|-------------------------------------------------------|--------------------|--------------------------|---------------------------------|----------------|------------------------------------|----------------------------------|
| Native                           | 1.7            | 94.3                                                  | 0.054              | —                        | —                               | —              | —                                  | —                                |
| K <sub>2</sub> PtCl <sub>4</sub> | 2.0            | 97.9                                                  | 0.089              | 0.279                    | 4                               | 2.25           | 0.58                               | 0.96                             |
| OsO <sub>4</sub>                 | 2.4            | 90.0                                                  | 0.034              | 0.204                    | 1                               | 1.16           | 0.81                               | —                                |
| AuCl <sub>3</sub>                | 2.0            | 99.9                                                  | 0.066              | 0.188                    | 1                               | 1.00           | 0.86                               | —                                |
| Overall figure of merit 0.530    |                |                                                       |                    |                          |                                 |                |                                    |                                  |
| Number of non-hydrogen atoms     |                |                                                       | Resolution (Å)     |                          | No. of reflections working/free |                | $R_{\text{cryst}}/R_{\text{free}}$ |                                  |
| Prot                             | Wat            | Ion                                                   |                    |                          |                                 |                | R.m.s. deviations bonds/angles     |                                  |
| 279                              | 30             | 1 Ca <sup>2+</sup><br>1 SO <sub>4</sub> <sup>2-</sup> | 15.0–1.7           |                          | 2732/227 (7.67%)                |                | 0.209/0.240                        |                                  |
|                                  |                |                                                       |                    |                          |                                 |                | 0.013 Å/2.498°                     |                                  |

The LR5 model that was built into the MIR electron density map contains residues 4–40 of the domain (Fig. 1); residues 1–3 were not visible in the MIR map or in 2*F*<sub>o</sub> – *F*<sub>c</sub> maps calculated with refined models of residues 4–40. The model also contains one highly coordinated calcium ion, 30 water molecules and one sulphate ion.

<sup>\*</sup>*R*<sub>sym</sub> = Σ ||*I*<sub>j</sub> – ⟨*I*⟩|/Σ *I*<sub>j</sub>, where *I*<sub>j</sub> is the intensity measurement for reflection *j* and ⟨*I*⟩ is the mean intensity for multiply recorded and symmetry-related reflections.

<sup>†</sup>*R*<sub>iso</sub> = Σ ||*F*<sub>ph</sub> – *F*<sub>p</sub>|/Σ ||*F*<sub>ph</sub> + *F*<sub>p</sub>|, where *F*<sub>ph</sub> and *F*<sub>p</sub> are the derivative and native structure factors, respectively.

<sup>‡</sup>Phasing power = ⟨*F*<sub>h</sub>⟩/⟨*E*⟩, where ⟨*F*<sub>h</sub>⟩ is the root-mean-square heavy-atom structure factor and *E* is the residual lack of closure error.

<sup>§</sup>Cullis *R* = Σ ||*F*<sub>ph</sub> ± *F*<sub>p</sub> – *F*<sub>h,c</sub>|/Σ ||*F*<sub>ph</sub> ± *F*<sub>p</sub>|, where *F*<sub>h,c</sub> is the calculated heavy-atom structure factor.

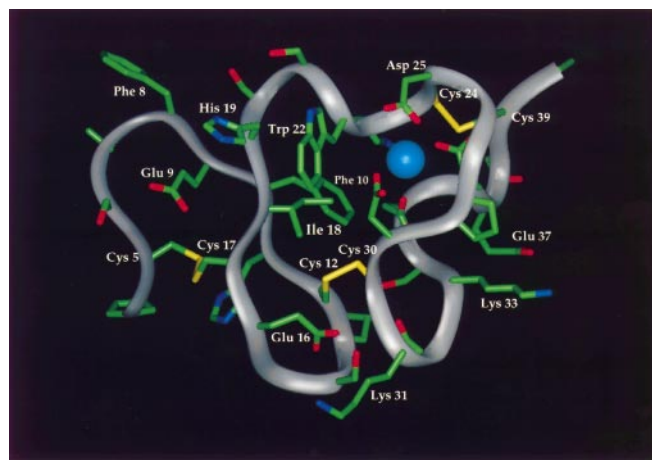
||Cullis *R*<sub>anomalous</sub> = Σ ||*F*<sub>ph</sub>(+) – *F*<sub>ph</sub>(–)| – 2*F*<sub>h</sub> sin α<sub>p</sub>|/Σ ||*F*<sub>ph</sub>(+) – *F*<sub>ph</sub>(–)|, where α<sub>p</sub> is the calculated protein phase.

*R*<sub>cryst,free</sub> = Σ ||*F*<sub>obs</sub> – *F*<sub>calc</sub>||/Σ ||*F*<sub>obs</sub>||, where the crystallographic and free *R* factors are calculated using the working and free reflection sets, respectively.

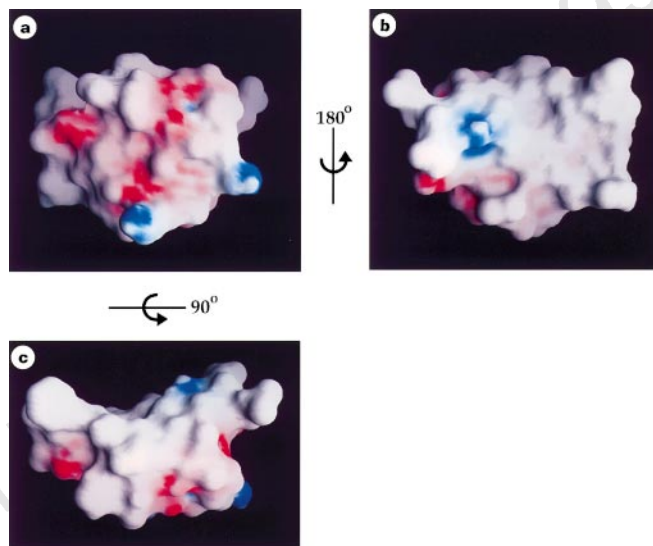
with the backbone amide of residue 14 and forms an Asx turn<sup>11</sup> by hydrogen bonding to the backbone amide of the conserved Ser at position 34. The use of both carboxylate oxygens explains why the relatively conservative Asp-to-Asn mutation at this position results in FH. Finally, Ser 34 is apparently conserved in LDL-A modules to act as a hydrogen-bond acceptor from the backbone amide of residue 13, and as a donor to the side chain of Asp 32. A proline at this position, in addition to kinking the chain near Ca<sup>2+</sup> ligands,

lacks side-chain hydrogen-bonding potential and the backbone amide to participate in the Asx turn with Asp 32. Not surprisingly, this mutation results in FH.

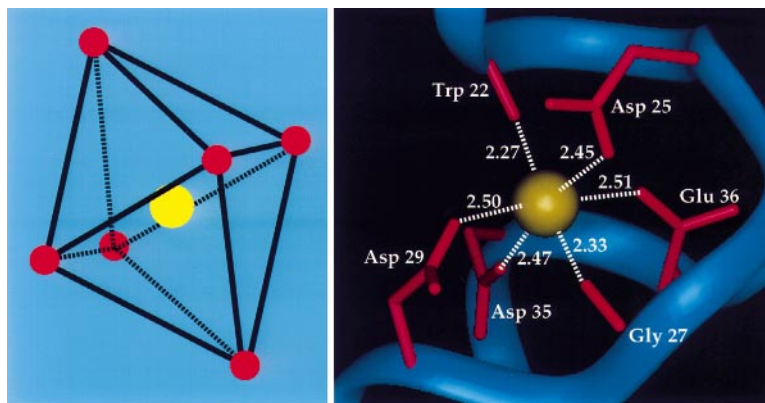
In addition to justifying the effects of FH mutations, the structure of LR5 calls into question existing models for normal LDLR function. These models postulate that the interaction between LDLR and its target occurs between the conserved acidic motif in the LDLR ligand-binding modules and positively charged residues



**Figure 2** Structure of LR5. Side chains are superimposed on a ribbon trace of the module backbone. Cysteine and additional residues are numbered as reference points. The Ca<sup>2+</sup> ion is indicated by a blue sphere. FH mutations are located at structurally significant residues (see text). The figure was generated using Insight II (Biosym Technologies). The LR5 structure has uniformly low *B*-factors for all residues after residue 4 (average main-chain *B*-factor = 10.7 Å<sup>2</sup>; average side-chain *B*-factor = 11.0 Å<sup>2</sup> for residues 5–40).



**Figure 4** Surface contours and charge density. **a**, This orientation of LR5 is similar to that in Fig. 2. The surface-exposed acidic residue on the left is Glu 9. The two regions of basic potential arise from Lys 31 and Lys 33. Glu 16 is above Lys 31 near the centre of this face. The Ca<sup>2+</sup> binding site region, towards the upper right, contributes little to the surface charge. **b**, The hydrophobic face of the molecule is viewed after a rotation of 180° around the vertical. **c**, A 90° rotation around the horizontal from the orientation in **a** generates this top view of LR5. The hydrophobic concave face is towards the top of the figure. Regions of basic potential ( $>16k_B T/e$ , where  $k_B$  is Boltzmann's constant and  $T$  is the absolute temperature) are shown in blue; acidic regions ( $<-14k_B T/e$ ) are in red. The figure was generated using GRASP<sup>27</sup>.



**Figure 3** LR5 Ca<sup>2+</sup> coordination compared to ideal octahedral coordination. Red balls indicate positions of coordinating oxygen atoms; the yellow ball shows the position of the Ca<sup>2+</sup> ion. Lines are drawn between coordinating atoms in the schematic diagram to show how a six-ligand coordination geometry generates an octahedron. Side chains involved in Ca<sup>2+</sup> coordination are superimposed on a ribbon trace of the LR5 backbone. Distances in Ångströms between coordinating oxygens and the Ca<sup>2+</sup> ion are indicated. The figure was generated using Ribbons<sup>26</sup>. The distances measured in the refined LR5 structure between the

side-chain carboxylate groups and the Ca<sup>2+</sup> ion are  $2.48 \pm 0.03$  Å. The distances between the backbone carbonyl oxygens and the Ca<sup>2+</sup> ion are  $2.30 \pm 0.03$  Å. The r.m.s.d. for the six oxygen atoms coordinating the Ca<sup>2+</sup> ion, as compared to an ideal octahedron with centre-to-vertex distances of 2.40 Å, is 0.21 Å, comparable to deviations seen for Ca<sup>2+</sup> sites in other proteins<sup>26</sup>. The octahedral coordination cages the Ca<sup>2+</sup> ion such that the folded domain structure is likely to inhibit exchange of the Ca<sup>2+</sup> ion with solvent, consistent with the ~70 nM Ca<sup>2+</sup>-binding affinity of the domain<sup>9</sup>.

of apolipoproteins<sup>2,12–14</sup>. However, our results show that the residues in the conserved acidic motif of LR5 are buried to participate in Ca<sup>2+</sup> coordination, instead of being exposed on the surface of the domain (Fig. 4a). Although lipoprotein uptake by members of the LDLR family may involve an electrostatic component, perhaps through association of the lipoproteins with cell-surface proteoglycans<sup>15</sup>, the LR5 structure demonstrates that the primary role for the conserved acidic residues in LDL-A modules is structural. It has been noted previously that only lipid-associated apolipoproteins bind with high affinity to the LDLR<sup>16</sup>; the LR5 structure suggests an alternative model for LDLR function in which a hydrophobic concave face (Fig. 4b, c) provides a lipoprotein-binding surface. □

## Methods

**Protein crystallization.** Protein was expressed, purified and refolded as described<sup>9</sup>. Protein stock solution contained 10 mg ml<sup>-1</sup> LR5 and 5 mM CaCl<sub>2</sub>. Crystals of LR5 were grown by the hanging-drop method at 15°C by equilibrating over a reservoir containing 2.1 M ammonium sulphate, adjusted to pH 5.0, and 25% sucrose. These crystals belong to the rhombohedral space group R3 ( $a = b = c = 32.04$  Å,  $\alpha = \beta = \gamma = 112.57^\circ$ ), with a single LR5 molecule in the asymmetric unit. Solvent content in the crystals is 28%.

**Data collection, structure determination and structure refinement.** Diffraction data were collected using a rotating anode source (Rigaku RU200) with an RAXIS IIC detector. All data were collected at -180°C using an X-stream cryogenic crystal cooler (MSC). Crystals were mounted in a 20 µm rayon fibre loop and flash-frozen in the gaseous nitrogen stream. Reflections were indexed and integrated with the program DENZO<sup>17</sup> and were scaled using SCALEPACK<sup>17</sup>. Subsequent data manipulations were carried out using the CCP4 package<sup>18</sup>.

Heavy-atom derivatization was accomplished by collecting crystals into reservoir solution, adding solid heavy atom to 1 mM, and soaking overnight. Heavy-atom sites were located for each derivative by Patterson methods. Heavy-atom refinement and phasing were carried out with MLPHARE (CCP4). Density modification was done on the MIR map using the program DM<sup>19</sup>. The 3-Å density-modified map was displayed with the program O (ref. 20) for tracing of the polypeptide chain, using the NMR structures of rLB1 and rLB2 (refs 7, 8) as a guide. Phases from the initial polyalanine model were combined with MIR phases using SIGMAA weighting<sup>21</sup>; side chains were readily built into the phase-combined map. A set of reflections (~7%) were held aside before refinement to monitor a free R-factor<sup>22</sup>. The model was refined against the remaining data to 1.7 Å with the program XPLOR<sup>23</sup>. B-factor refinement was done first, followed by simulated annealing. A bulk solvent correction was then applied and ordered waters were added.

Received 26 March; accepted 19 May 1997.

1. Goldstein, J. L. *et al.* Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Annu. Rev. Cell Biol.* **1**, 1–39 (1985).
2. Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).
3. Esser, V. *et al.* Mutational analysis of the ligand binding domain of the low density lipoprotein receptor. *J. Biol. Chem.* **263**, 13282–13290 (1988).
4. Russell, D. W., Brown, M. S. & Goldstein, J. L. Different combinations of cysteine-rich repeats mediate binding of low density lipoprotein receptor to two different proteins. *J. Biol. Chem.* **264**, 21782–21688 (1989).
5. Altschul, S. F. *et al.* Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
6. Goldstein, J. L., Hobbs, H. H. & Brown, M. S. in *The Metabolic and Molecular Bases of Inherited Disease* (eds Scriver, C. S., Beaudet, A. L., Sly, W. S. & Valle, D.) 1981–2030 (McGraw Hill, New York, 1995).
7. Daly, N. L. *et al.* Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. *Proc. Natl Acad. Sci. USA* **92**, 6334–6338 (1995).
8. Daly, N. L., Djordjevic, J. T., Kroon, P. A. & Smith, R. Three-dimensional structure of the second cysteine-rich repeat from the human low-density lipoprotein receptor. *Biochemistry* **34**, 14474–14481 (1995).
9. Blacklow, S. C. & Kim, P. S. Protein folding and calcium binding defects arising from familial hypercholesterolemia mutations of the LDL receptor. *Nature Struct. Biol.* **3**, 758–762 (1996).
10. van Driel, I. R., Goldstein, J. L., Südhof, T. C. & Brown, M. S. First cysteine-rich repeat in ligand-binding domain of low density lipoprotein receptor binds Ca<sup>2+</sup> and monoclonal antibodies, but not lipoproteins. *J. Biol. Chem.* **262**, 17443–17449 (1987).
11. Abbadi, A. *et al.* Involvement of side functions in peptide structures: the Asx turn. Occurrence and conformational aspects. *J. Am. Chem. Soc.* **113**, 2729–2735 (1991).
12. Südhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* **228**, 815–822 (1985).
13. Mahley, R. W. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* **240**, 622–630 (1988).
14. Wilson, C. *et al.* Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. *Science* **252**, 1817–1822 (1991).
15. Ji, Z.-S., Fazio, S., Lee, Y.-L. & Mahley, R. W. Secretion-capture role for apolipoprotein E in remnant

lipoprotein metabolism involving cell surface heparan sulfate proteoglycans. *J. Biol. Chem.* **269**, 2764–2772 (1994).

16. Innerarity, T. L., Pitas, R. E. & Mahley, R. W. Binding of arginine-rich (E) apolipoprotein after recombination with phospholipid vesicles to the low density lipoprotein receptors of fibroblasts. *J. Biol. Chem.* **254**, 4186–4190 (1979).
17. Otwinowski, Z. & Minor, W. *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, UK, 1993).
18. CCP4. The SERC (UK) Collaborative Computing Project No. 4 *A Suite of Programs for Protein Crystallography* (SERC Daresbury Laboratory, Warrington, UK, 1979).
19. Cowtan, K. D. Joint CCP4 and ESRF-ACBM Newsletter on Protein Crystallography **31**, 34–38 (1994).
20. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
21. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
22. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
23. Brünger, A. T. *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
24. Yamamoto, T. *et al.* The human LDL receptor: a cysteine-rich protein with multiple *Alu* sequences in its mRNA. *Cell* **39**, 27–38 (1984).
25. Carson, M. & Bugg, C. E. Algorithm for ribbon models of proteins. *J. Mol. Graph.* **4**, 121–122 (1986).
26. McPhalen, C. A., Strynadka, N. C. J. & James, M. N. G. Calcium binding sites in proteins: a structural perspective. *Adv. Prot. Chem.* **42**, 77–144 (1991).
27. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

**Acknowledgements.** We thank S. J. Gamblin for technical advice and discussions. J.M.B. is a Whitehead Fellow and acknowledges support from the W. M. Keck Foundation. This research utilized the W. M. Keck Foundation X-ray Crystallography Facility at the Whitehead Institute, and was supported by an NIH grant (to R. D. Rosenberg) from the Program of Excellence in Molecular Biology.

Correspondence and requests for materials should be addressed to P.S.K. Coordinates have been deposited in Brookhaven Protein Data Bank under accession code 1AJJ.

## Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP

Katrin Rittinger<sup>\*†</sup>, Philip A. Walker<sup>\*†</sup>, John F. Eccleston<sup>\*</sup>, Kurshid Nurmahomed<sup>\*</sup>, Darerca Owen<sup>‡</sup>, Ernest Laue<sup>‡</sup>, Steven J. Gamblin<sup>\*</sup> & Stephen J. Smerdon<sup>\*</sup>

<sup>\*</sup> National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

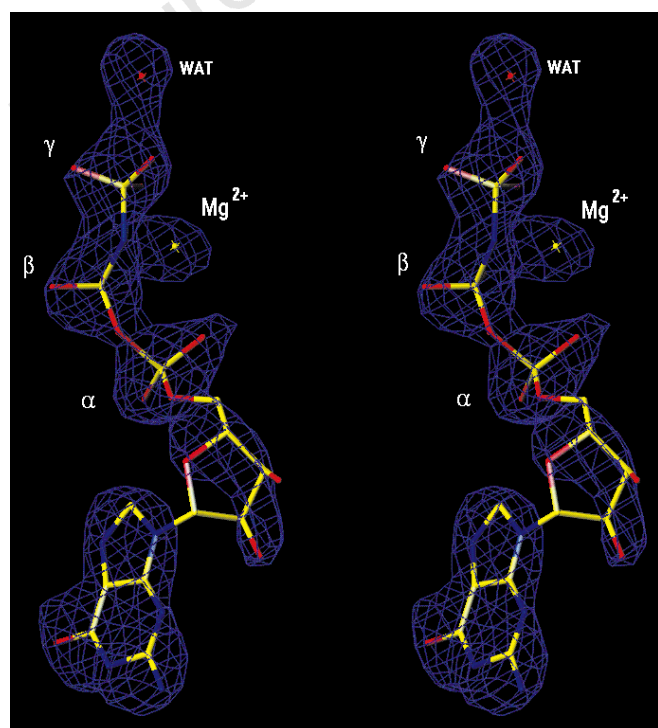
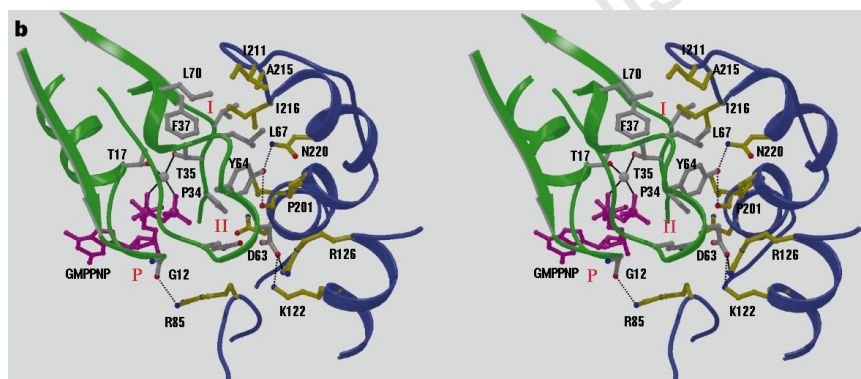
<sup>‡</sup> Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

<sup>†</sup> These authors contributed equally to this work.

Small G proteins transduce signals from plasma-membrane receptors to control a wide range of cellular functions<sup>1,2</sup>. These proteins are clustered into distinct families but all act as molecular switches, active in their GTP-bound form but inactive when GDP-bound. The Rho family of G proteins, which includes Cdc42Hs, activate effectors involved in the regulation of cytoskeleton formation, cell proliferation and the JNK signalling pathway<sup>3–9</sup>. G proteins generally have a low intrinsic GTPase hydrolytic activity but there are family-specific groups of GTPase-activating proteins (GAPs) that enhance the rate of GTP hydrolysis by up to 10<sup>5</sup> times<sup>10,11</sup>. We report here the crystal structure of Cdc42Hs, with the non-hydrolysable GTP analogue GMPPNP, in complex with the GAP domain of p50rhoGAP at 2.7 Å resolution. In the complex Cdc42Hs interacts, mainly through its switch I and II regions, with a shallow pocket on rhoGAP which is lined with conserved residues. Arg 85 of rhoGAP interacts with the P-loop of Cdc42Hs, but from biochemical data and by analogy with the G-protein subunit G<sub>iα1</sub> (ref. 12), we propose that it adopts a different conformation during the catalytic cycle which enables it to stabilize the transition state of the GTP-hydrolysis reaction.



**Figure 1 a**, The structure of the Cdc42Hs-GMPPNP/p50rhoGAP complex viewed along the protein-protein interface. Cdc42Hs is coloured green, with the conserved switch I (residues 32 to 40), switch II (residues 60 to 67) and P-loop (residues 10 to 17) regions in red. These bind to a shallow depression in rhoGAP (blue) formed by the A–Al loop and helices B and F which are highlighted in magenta. Formation of the complex buries 1,807 Å<sup>2</sup> of accessible surface, of which 61% (1,115 Å<sup>2</sup>) is made up of non-polar side-chain atoms. The Cα positions of Arg 85, and Tyr 64, are highlighted. **b**, Stereo view of the interactions between Cdc42Hs and p50rhoGAP viewed from the same direction as in **a**. Secondary-structure elements of Cdc42Hs and rhoGAP are coloured green and blue respectively, with Cdc42Hs side chains in grey and those of rhoGAP in yellow. In both cases, oxygen and nitrogen atoms are red and blue respectively. The electron density for residues 29–31 of the switch I region of Cdc42Hs is poor. Electron density for Tyr 32<sub>c</sub> is relatively poor at the current state of refinement, but it does refine adequately in a conformation which makes additional contacts with Lys 189, Thr 191, and Asn 194, Arg 85, occupies a conformation close to the P-loop of Cdc42Hs and interacts either through direct hydrogen bonding with the C=O of Gly 12<sub>c</sub> or through water mediated contacts. Gln61<sub>c</sub>, which is important in the GTP hydrolysis reaction, is disordered in this structure as it is in X-ray structures of Ras. (Panels **a** and **b** were produced using MOLSCRIPT<sup>29</sup>.)



**Figure 2** Electron density omit map for the bound GMPPNP-Mg. The  $|F_o| - |F_c|$  electron density map was generated from coefficients calculated by removing GMPPNP-Mg and WAT from the atomic model for 10 cycles of refinement before 3-fold averaging the electron density map according to the non-crystallographic symmetry. The electron density map is contoured at 3.0 standard deviations above the average density and clearly shows the presence of the Mg<sup>2+</sup> ion which is coordinated to β- and γ-phosphate oxygen atoms and a water molecules which is tightly associated with the terminal phosphate group of the GMPPNP. This water is structurally conserved in small GTPases and in the transducin family and is the presumed nucleophile for attack of the γ-phosphate during GTP hydrolysis.



**Table 1 Summary of crystallographic analysis**

| Crystal resolution (Å)              | $R_{\text{sym}}$ (%) | Completeness (%)<br>$N_{\text{total}}/N_{\text{unique}}$ |                   | $\alpha$                          | $\beta$                          | $\gamma$               | t1                     | t2                      | t3    | c.c.* | $R_{\text{cyst}}^{\dagger}$ |
|-------------------------------------|----------------------|----------------------------------------------------------|-------------------|-----------------------------------|----------------------------------|------------------------|------------------------|-------------------------|-------|-------|-----------------------------|
| R3                                  | 7.7 (34.0)           | 91.4 (55.0)                                              | GAP               | 14.73                             | 85.23                            | 212.7                  | 0.802                  | 0.058                   | 0.709 | 38.1  | 38.8                        |
| 15–3.2                              |                      | 18,552/9,250                                             | Cdc42Hs           | 85.16                             | 102.19                           | 69.65                  | 0.989                  | 0.025                   | 0.253 |       |                             |
| P1                                  | 6.4 (24.7)           | 97.6 (96.5)                                              | (1)               | 272.6                             | 30.6                             | 38.3                   | 0.000                  | 0.000                   | 0.000 |       |                             |
| 25–2.7                              |                      | 246,060/49,357                                           | (2)               | 239.4                             | 91.3                             | 310.5                  | 0.416                  | 0.088                   | 0.505 |       |                             |
|                                     |                      |                                                          | (3)               | 178.5                             | 59.4                             | 221.2                  | 0.505                  | 0.582                   | 0.087 | 42.1  | 37.2                        |
| Averaging and refinement statistics |                      |                                                          |                   |                                   |                                  |                        |                        |                         |       |       |                             |
| Resolution (Å)                      | Averaging c.c.*      | Atoms                                                    | Solvent molecules | $R_{\text{crist}}(\%)^{\ddagger}$ | $R_{\text{free}}(\%)^{\ddagger}$ | R.m.s. ncs (Å) $^{\S}$ | R.m.s. bond length (Å) | R.m.s. bond angle (deg) |       |       |                             |
| 12–2.7                              | 0.928                | 8,907                                                    | 310               | 23.1                              | 28.4                             | 0.06                   | 0.014                  | 2.7                     |       |       |                             |

Figures in brackets refer to the outermost resolution shell.  $R_{\text{merge}} = \sum_i (|I_i| - \langle I \rangle) / \sum_i I_i$ , where  $I_i$  is the observed intensity and  $\langle I \rangle$  is the average intensity.  $\alpha$ ,  $\beta$ ,  $\gamma$ , t1, t2 and t3 are the eulerian rotation angles and fractional translational shifts as defined in the AMORE package. (1), (2) and (3) refer to each of the three complexes present in the crystallographic asymmetric unit.

\* c.c., Correlation coefficient; c.c. =  $\sum (F_o - |F_o|) \times (|F_{\text{prcalc}}| - |F_{\text{prcalc}}|) / \sum (|F_o| + |F_{\text{prcalc}}|)$ .

$R_{\text{crist}} = \sum (|F_o - F_{\text{prcalc}}|) / \sum F_o$  for all reflections, where  $F_o$  and  $F_{\text{prcalc}}$  are the observed and calculated structure amplitudes respectively.

$R_{\text{free}}$  is the same as  $R_{\text{crist}}$ , but calculated on the 5% of data excluded from refinement.

$^{\S}$  R.m.s. difference between non-crystallographic symmetry equivalents on all atoms.

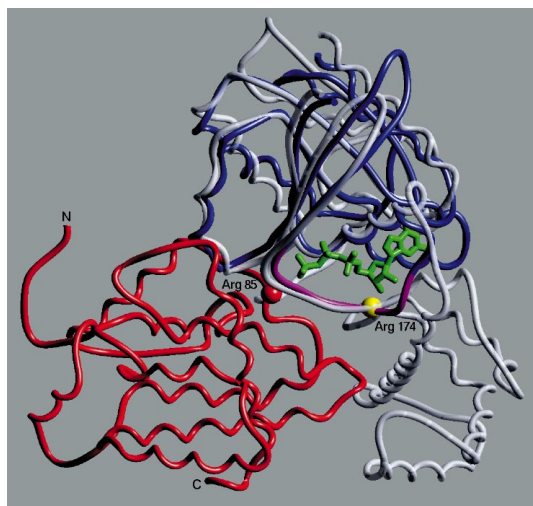
The crystal structures of Ras and Rho proteins are similar<sup>13–16</sup>, and of the 41 common residues, most are located around the guanine-nucleotide-binding site. It is likely, therefore, that the GTP-hydrolysis switch operates in the same way in both families. Sequence conservation within, but not between, the Rho and Ras families reflects the fact that they activate different targets. By contrast, there is no detectable tertiary structure or sequence similarity between the largely helical GAP domains that act on Rho family proteins and those that act on Ras proteins<sup>17,18</sup>. The heterotrimeric G-protein family share structural and functional parallels with small G proteins. Members of this family, which include transducin and the closely related G-protein subunit  $G_{i\alpha 1}$ , are important in physiological responses to neurotransmitters, hormones and sensory stimuli<sup>12,19,20</sup>. They too act as GTP-modulated molecular switches and contain a Ras-like GTPase domain. Heterotrimeric G proteins also possess, an extra, largely helical, domain that is important for enhancing their intrinsic GTPase rate, which is  $10^2$  fold faster than for Ras proteins. In addition, there are GAPs specific for heterotrimeric G proteins. The structure determination of RGS4, a  $G_{i\alpha 1}$ -specific GAP, again reveals it to be both largely helical and of a different tertiary structure to either Ras- or Rho-specific GAPs<sup>21</sup>. It is intriguing that for the three different families of GAPs whose structures are known, all are mostly helical but share no sequence or tertiary structure similarity. They all appear to perform the same function, but it remains to be seen whether they share a related mechanism of action.

The crystal structure of the Cdc42Hs-GMPPNP/p50rhoGAP complex was solved by molecular replacement using crystallographic models for both the G protein<sup>16</sup> and GAP domain<sup>18</sup>. The initial electron density maps at 2.7 Å were of good quality because of the availability of well refined models for both components, the fact that there were only limited rearrangements of the two starting models, and the presence of non-crystallographic symmetry. An example of the averaged electron density map where the model has not been included in the calculation of starting phases is shown in Fig. 2. The current resolution of 2.7 Å does not enable us to make detailed mechanistic proposals for GAP-enhanced GTPase activity, but it does allow us to identify the key residues and structures involved in the formation of the complex.

The structure of the complex shows that the Cdc42Hs contacts with rhoGAP are mainly mediated through its switch I and II regions and the P-loop (Fig. 1a). Together, these regions contact residues of rhoGAP that are primarily located in and around a

shallow pocket formed between the B and F helices, the A–Al loop and the F–G loop. As described previously, there is a preponderance of conserved residues associated with this rhoGAP pocket region<sup>18</sup>. The two proteins pack together so that the switch I and II regions of the Cdc42Hs slot into this pocket, making polar and apolar interactions with the rhoGAP domain (for ease of description, residues will be identified as belonging to Cdc42Hs and rhoGAP by using the subscripts 'c' and 'r' respectively). Contacts mediated through the switch II region include salt-bridging interactions between Asp 63<sub>c</sub> and Lys 122<sub>r</sub>, which is conserved in rhoGAP proteins, and Arg 126<sub>r</sub> on the B-helix. Tyr 64<sub>c</sub> hydrogen-bonds with both the side chain of Asn 220<sub>r</sub> on the G-helix and the main-chain C=O of Val 197<sub>r</sub>, whereas Leu 67<sub>c</sub> packs against the F–G loop (Fig. 1b). Although structural data are only available for the GMPPNP form of a Rho protein<sup>16</sup>, it is interesting that the switch II region in Ras, including Tyr 64, has been implicated in conformational rearrangements following GTP hydrolysis<sup>22,23</sup>. Such a conformational change in Cdc42Hs would disrupt many of the interactions with p50rhoGAP observed in this complex structure. In common with the structure of Rac1-GMPPNP, three residues (29–31) of the switch I region are not ordered in this complex. However, the remainder of this loop is well ordered and contributes significantly to the nonpolar interactions at the molecular interface (Fig. 1b). In particular, residues Val 36<sub>c</sub> and Phe 37<sub>c</sub>, towards the end of switch I, make hydrophobic packing interactions with residues in the F–G loop of rhoGAP. This probably accounts for the F–G loop being ordered in this complex, but not in crystals of p50rhoGAP alone. As residues in the F–G loop of the rhoGAPs are not conserved, it seems that these interactions contribute stabilization, rather than specificity, to complex formation.

We have previously argued that Arg 85<sub>r</sub> is likely to be important in G-protein interaction<sup>18</sup>. In this complex, Arg 85<sub>r</sub> makes hydrogen-bonding interactions with the main-chain carbonyl of the conserved Gly 12<sub>c</sub> on the P-loop. However, the role of this residue in rhoGAP-activated hydrolysis of GTP may well be similar to the roles of arginines in transducin,  $G_{i\alpha 1}$  and Ras/rasGAP. Transducin-GDP and  $G_{i\alpha 1}$ -GDP both bind to  $\text{AlF}_4^-$ , which mimics the terminal phosphate of GTP. In contrast Ras-GDP can only form this so-called transition-state analogue in the presence of p120GAP or neurofibromin<sup>24</sup>. This hypothesis is supported by both biochemical and structural data. Changes in fluorescence anisotropy have shown that rhoGAP binds tightly to rhoA-mantGDP (where mantGDP is fluorescently labelled 2'(3')-O-(N-methylanthraniloyl)GDP) in the presence, but not in



**Figure 3** Comparison of the Cdc42Hs-GMPPNP/p50rhoGAP complex with the structure of transducin- $\alpha$ -GTP $\gamma$ S. Here, Cdc42Hs is shown in blue, with rhoGAP in red and transducin- $\alpha$  in white. Transducin- $\alpha$  consists of two domains: one is structurally homologous to the small GTPase family and the other is a smaller  $\alpha$ -helical domain unique to the heterotrimeric G-protein family. Least-squares overlap<sup>28</sup> of Cdc42Hs with the GTPase domain of transducin gives an r.m.s. fit of 1.7 Å for 122 structurally equivalent C $\alpha$  atoms. Arg 174 has been proposed to be involved in transition-state stabilization during GTP hydrolysis in transducin. Its C $\alpha$  position is shown in yellow and is positioned within a piece of extended structure corresponding to the effector or switch I region of Cdc42Hs (magenta). In contrast, Arg 85, which appears to fulfil a similar role in GTPase activation by p50rhoGAP, is situated on the A–Al loop on the opposing face of the triphosphate moiety of the nucleotide analogue.

the absence, of AlF<sub>4</sub><sup>−</sup>. Furthermore, mutation of Arg 85<sub>r</sub> to lysine abolishes the formation of the rhoGAP/rhoA-GDP-AlF<sub>4</sub> complex and reduces the rhoGAP activation of GTP hydrolysis by a factor of  $\sim 10^4$  (D. Graham, P. N. Lowe, K. R. and J. F. E., unpublished results). These data suggest that Arg 85<sub>r</sub> interacts with the GDP-AlF<sub>4</sub><sup>−</sup> transition state analogue. An alternative side-chain rotamer of Arg 85<sub>r</sub> to that observed in the current model would enable this interaction to occur. In this context, when G<sub>iα1</sub> is complexed with GTP- $\gamma$ S, Arg 178 is not well ordered but does interact with the transition state analogue in the complex with GDP-AlF<sub>4</sub><sup>−</sup>. However, in the case of transducin, the equivalent residue Arg 174 is ordered in both types of guanine-nucleotide-analogue complex<sup>19,20</sup>.

Figure 3 shows the Cdc42Hs-GMPPNP/rhoGAP complex and transducin-GDP-AlF<sub>4</sub><sup>−</sup> structures in which the cores of the Cdc42Hs and Ras-like domain of transducin have been aligned. Two points emerge from this comparison. First, although the non-Ras-like domain of transducin and the GAP domain of the complex are both extensively helical, they have entirely different folds and occupy exclusively different positions with respect to the G-protein component. Second, the C $\alpha$  positions of Arg 174 of transducin and Arg 85 of rhoGAP, both of which apparently bind to the GDP-AlF<sub>4</sub><sup>−</sup> transition-state analogue, are 14 Å apart. Nevertheless, the guanidinium moiety of Arg 174 is in contact with the nucleotide analogue in the crystal structure of the transducin-GDP-AlF<sub>4</sub><sup>−</sup> complex, whereas in our GMPPNP complex, a simple side-chain rotation of Arg 85 would be sufficient to bring its guanidinium group into a similar position and fulfil the same biochemical function.

The structure of RGS4 bound to AlF<sub>4</sub><sup>−</sup>-activated G<sub>iα1</sub> has been solved<sup>21</sup> and shows that this GAP domain is composed of nine  $\alpha$ -helices arranged into two subdomains. The RGS4 makes extensive contacts with the switch I, II and III regions of the Ras-like domain of G<sub>iα1</sub>, but very limited contacts with the  $\alpha$ -helical domain. Not surprisingly, the RGS4 occupies the same overall position as rhoGAP does in the superposition shown in Fig. 3. However, rhoGAP and RGS4 have quite different helical folds and RGS4 does not contribute catalytic residues for the GTP-hydrolysis reaction. Instead, it stabilizes the switch regions of G<sub>iα1</sub>, thus enhancing the intrinsic GTP-hydrolysis reaction. Therefore, although rhoGAP and RGS4 both interact with their respective G proteins through the switch regions, the contacts are quite different in detail. It is as though rhoGAP combines the function of RGS4 and the G<sub>iα1</sub> helical domain in one protein.

It was previously uncertain whether GAP acts allosterically, or if it participates directly in catalysis. Our results suggest that Arg 85<sub>r</sub> is important in transition-state stabilization, which is in keeping with similar roles ascribed to specific arginine residues in Ras/rasGAP

and heterotrimeric G proteins. Nonetheless, rhoGAP still enhances GTPase activity by 30-fold when this arginine is mutated to lysine. It is likely that this remaining activation is provided by rhoGAP stabilizing the switch I and II regions of the G protein in a manner similar to that suggested for RGS4 acting on G<sub>iα1</sub>. A docking model has been suggested for Ras/p120rasGAP (ref. 17); if this model proves to be correct then a detailed comparison of these two systems will provide insight into the apparently similar catalytic chemistry achieved by these two different GAP structures. □

#### Methods

**Crystallization.** Cdc42Hs (fragment 1–184) was cloned via pBluescript (Stratagene) into the expression vector pGEX-4T1 and the protein purified by standard protocols<sup>25</sup>, the p50rhoGAP (fragment 198–439) was expressed and purified as described previously<sup>10</sup>. Crystals were grown by vapour diffusion at 4 °C from 20% PEG 4000, 100 mM sodium cacodylate, pH 5.9, 10 mM MgCl<sub>2</sub>, 3 mM DTT, 3 mM NaN<sub>3</sub>. Initial crystals belonged to space group R3 ( $a = b = c = 78.2$  Å,  $\alpha = \beta = \gamma = 90.1^\circ$ ), with one Cdc42Hs/rhoGAP complex in the crystallographic asymmetric unit. Better diffracting crystals were subsequently obtained from similar conditions which belonged to space group P1 with three complexes in the unit cell ( $a = 78.0$  Å,  $b = 78.1$  Å,  $c = 78.3$  Å,  $\alpha = 90.1^\circ$ ,  $\beta = 90.2^\circ$ ,  $\gamma = 90.3^\circ$ ).

**Data collection and processing.** Diffraction datasets were collected from single crystals at 100 K using an Oxford Cryosystems device. Crystals were prepared for flash cooling by immersion in mother liquor made up to 15% PEG 400. Native datasets for both crystals were collected on an Raxis II image plate mounted on a Rigaku RU200 generator. Datasets for the P1 crystal form were also collected at station 5.2R at Trieste and BM14 at ESRF but were subsequently superseded by a largely complete and highly redundant dataset collected at Daresbury 7.2 on a large MAR image detector (statistics for which are shown in Table 1). Data were processed with DENZO<sup>26</sup> and SCALEPACK<sup>26</sup>; all subsequent calculations being performed with the CCP4<sup>27</sup> program suite.

**Structure solution and refinement.** Initial phases for the complex were calculated by molecular replacement. The starting models were the p50rhoGAP fragment structure (Brookhaven accession code 1rgp) and the Rac1 homologue of Cdc42 (accession code 1mh1). Using the AMoRe<sup>27</sup> package, rotation/translation parameters were calculated for the R3 dataset using all data between 15 and 3.2 Å. After rigid-body refinement, the p50rhoGAP molecular replacement produced a correlation coefficient of 0.26; the next highest peak was 0.18. The rhoGAP fragment was then fixed while translation function searches were carried out with the Rac1 model. The molecular replacement solution with both p50rhoGAP and Rac1 gave a correlation coefficient of 0.38, with the next highest peak being 0.26. The same unique solution was found if the molecular replacement was carried out in the reverse order: first determining and then fixing the parameters for the Rac1 model, followed by translation searches with the rhoGAP model. The two-component model for the complex (without

refinement) was then used for molecular replacement calculations against the P1 dataset. Three solutions were found, related by a local three-fold axis approximately aligned with the body-diagonal of the unit cell, producing a correlation coefficient of 0.42 (the next highest solution being 0.28). Electron density for parts of the complex missing from the starting model coordinates, including some residues at the C terminus and in the F–G loop of p50rhoGAP, were readily interpretable. Electron density maps were 3-fold averaged using the program DM<sup>27</sup>. The subset of reflections flagged for omission from the refinement process were also omitted from DM phase refinement. Manual model adjustment was carried out using the O program<sup>28</sup>; reciprocal space refinement was done with REFMAC<sup>27</sup> using DM-modified phases as additional restraints. Non-crystallographic positional and thermal restraints were applied in PROTEIN/REFMAC, with tight targets for both main and side-chains. Water molecules were built into peaks greater than 3.5 $\sigma$  in averaged difference electron density maps using the program ARP<sup>27</sup> and retained only if they met strict B-factor (less than 1.5 greater than average for protein) and stereochemical requirements.

Received 25 April; accepted 25 June 1997.

1. Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125–132 (1990).
2. Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**, 117–127 (1991).
3. Paterson, H. F. *et al.* Microinjection of recombinant p21rho induces rapid changes in cell morphology. *J. Cell Biol.* **111**, 1001–1007 (1990).
4. Hall, A. The cellular functions of small GTP-binding proteins. *Science* **249**, 635–640 (1990).
5. Ridley, A. J. & Hall, A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389–399 (1992).
6. Qiu, R.-G., Chen, J., Kirn, D., McCormick, F. & Symns, M. An essential role for Rac in Ras transformation. *Nature* **374**, 457–459 (1995).
7. Ridley, A. J. Rho: theme and variations. *Curr. Biol.* **6**, 1256–1264 (1996).
8. Coso, O. A. *et al.* The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signaling pathway. *Cell* **81**, 1137–1146 (1995).
9. Minden, A., Lin, A., Claret, F.-X., Abo, A. & Karin, M. Selective activation of the JNK signaling cascade and c-Jun transcriptional activity by the small GTPases Rac and Cdc42Hs. *Cell* **81**, 1147–1157 (1995).
10. Lancaster, C. A. *et al.* Characterization of rhoGAP. *J. Biol. Chem.* **269**, 1137–1142 (1994).
11. Lamarche, N. & Hall, A. GAPs for rho-related GTPases. *Trends Genet.* **10**, 436–440 (1994).
12. Coleman, D. E. *et al.* Structures of active conformations of G<sub>12i</sub> and the mechanism of GTP hydrolysis. *Science* **265**, 1405–1412 (1994).
13. Pai, E. F. *et al.* Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**, 2351–2359 (1990).
14. Schlichting, I. *et al.* Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature* **345**, 309–315 (1990).
15. Prive, G. G. *et al.* X-ray crystal structures of transforming p21 ras mutants suggest a transition-state stabilization mechanism for GTP hydrolysis. *Proc. Natl Acad. Sci. USA* **89**, 3649–3653 (1992).
16. Hirschberg, M., Stockley, R. W., Dodson, G. & Webb, M. R. The crystal structure of human rac1, a member of the rho-family complexed with a GTP analogue. *Nature Struct. Biol.* **4**, 147–152 (1997).
17. Scheffzek, K., Lautwein, A., Kabsch, W., Ahmadian, M. R. & Wittinghofer, A. Crystal structure of the GTPase-activating domain of human p120GAP and implications for the interaction with Ras. *Nature* **384**, 591–596 (1996).
18. Barrett, T. *et al.* The structure of the GTPase-activating domain from p50rhoGAP. *Nature* **385**, 458–461 (1997).
19. Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 Å crystal structure of transducin- $\alpha$  complexed with GTP- $\gamma$ S. *Nature* **366**, 654–663 (1993).
20. Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. GTPase mechanism of G-proteins from the 1.7 Å crystal structure of transducin-GDP AlF<sub>4</sub>. *Nature* **372**, 276–279 (1994).
21. Tesmer, J. J. G., Berman, D. M., Gilman, A. G. & Sprang, S. R. Structure of RGS4 bound to AlF<sub>4</sub>-activated G<sub>12i</sub>: Stabilization of the transition state for GTP hydrolysis. *Cell* **89**, 251–261 (1997).
22. Milburn, M. V. *et al.* Molecular switch for signal transduction: Structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939–945 (1990).
23. Goody, R. S. *et al.* Studies on the structure and mechanism of H-ras p21. *Phil. Trans. R. Soc. Lond. B* **336**, 3–11 (1992).
24. Mittal, R., Ahmadian, M. R., Goody, R. S. & Wittinghofer, A. Formation of a transition-state analog of Ras GTPase reaction by Ras GDP, tetrafluoroaluminate, and GTPase-activating proteins. *Science* **273**, 115–117 (1996).
25. Smith, D. B. & Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31–40 (1988).
26. Otwinowski, Z. & Minor, W. *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, 1993).
27. CCP4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

**Acknowledgements.** We thank M. Hirschberg for making coordinates of Rac1 available before general release, E. Skordalakes for assistance with data collection, A. Hall for the gift of rhoA and rhoGAP expression systems, E. Manser and L. Lim for construction and characterization of the Cdc42Hs $\Delta$ 7 expression system, G. Dodson and D. Trentham for critical reading of this manuscript, and A. Savoia (Trieste) and A. Thompson (ESRF) for beamline assistance. D.O. and E.L. Thank the European commission for financial support and the BBSRC and Wellcome Trust for support to the Cambridge Centre for Molecular Recognition. K.R. is funded by an EEC TMR fellowship.

Correspondence and requests for materials should be addressed to S.J.G. (e-mail: s-gambli@anika.nimr.mrc.ac.uk). Coordinates have been deposited with the Brookhaven DataBank under accession number 1AM4.

## corrections

# A new chromosomal protein essential for mitotic spindle assembly

Jing-Ping Yeo, Frank Alderuccio & Ban-Hock Toh

*Nature* **367**, 288–291 (1994)

In this Letter, we identified a chromosomal antigen and its corresponding cDNA by using a human autoimmune serum. We named this 47K autoantigen RMSA1 (for regulator of mitotic spindle assembly 1). Tugendreich *et al.*<sup>1</sup> and Margalit *et al.*<sup>2</sup> observed that the open reading frame (ORF) we ascribed to the cDNA of RMSA1 contains two Alu repeat sequences. Although truncated Alu repeats have been identified in the ORFs of messenger RNAs, Tugendreich *et al.* suggested that our reported cDNA may be an expressed pseudogene or a cloning artefact. We have now repeated the experiments (but without J.-P.Y., who has left our laboratory) and find that the reported RMSA cDNA is indeed a cloning artefact and that the identity of the chromosomal protein, which we have confirmed reacts by immunofluorescence with the autoimmune serum, remains unknown. □

1. Tugendreich, S. *et al.* *Nature* **370**, 106 (1994).
2. Margalit, H., Nadir, E. & Ben-Sasson, S. A. *Cell* **78**, 173–174 (1994).

# RNA recognition and translational regulation by a homeodomain protein

Josh Dubnau & Gary Struhl

*Nature* **379**, 694–699 (1996)

In this Article, we presented genetic and biochemical evidence that Bicoid (Bcd) protein binds and regulates the translation of *caudal* (*cad*) messenger RNA through direct interactions between the Bcd homeodomain and a discrete target sequence in the 3' UTR of *cad* mRNA, the Bcd response element (BRE). Although our main findings have been verified by additional experiments (see page 634 in this issue)<sup>1</sup>, we cannot reproduce a subset of the results under the conditions initially reported.

The results we cannot repeat are: (1) the ultraviolet crosslinking and band-shift experiments which provided evidence for sequence-specific interactions between the Bcd homeodomain and the BRE (Figs 4c and 5); and (2) a bicistronic transcript experiment (Fig. 6) which provided evidence that the BRE mediates Bcd-dependent regulation of *Cad* at the level of translational initiation. However, biochemical evidence that the Bcd homeodomain can bind the BRE in a sequence-specific fashion has now been obtained in independent experiments using different binding conditions<sup>1</sup>. In addition, we have successfully repeated the key genetic experiments indicating that Bcd regulates *Cad* expression in a way that depends on both the Bcd homeodomain and the *cad* BRE (see Figs 1, 2c i, v and vi, 2d vi (bcd<sup>+</sup>, bcd<sup>E1</sup>), and 3 of our Article).

These errors do not alter our main interpretation that the homeodomain of Bcd mediates RNA recognition and translational regulation by the intact protein, but we apologize for any confusion they may have caused. □

1. Chan, S.-K. & Struhl, G. Sequence-specific RNA binding by Bicoid. *Nature* **388**, 634 (1997).

# Synapse specificity of long-term potentiation breaks down at short distances

Florian Engert & Tobias Bonhoeffer

*Nature* 388, 279–284 (1997).

Owing to a fault in the final production process, three of the ten panels in the left column of Fig. 3a have been greyed out. This might have conveyed the erroneous impression that three of the ten experiments were conducted differently. The correct figures are reprinted here to emphasize that this is not the case.

